

KM autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
 KM autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
 KM Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
 KM coagulation disorder; inflammatory disorder; Crohn's disease; incision;
 KM tissue regeneration; wound healing; burn; haematopoiesis;
 KM myeloid cell deficiency; lymphoid cell deficiency.
 OS Homo sapiens.
 FN WO200177288-A2.
 EN 18-OCT-2001.
 PD 23-MAR-2001; 2001WO-US10224.
 PF 06-APR-2000; 2000US-195582P.
 PR (GEMV) GENETICS INST INC.
 PA
 PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
 PI Gulluxota K, Graham JK;
 DR MPI; 2002-179321/23.
 XX Five hundred and ninety two polynucleotides derived from a variety of
 PT human tissue sources which encode secreted proteins, useful for
 PT treating immune deficiencies and disorders such as autoimmune disorders
 PT
 PS Claim 1; Page 78; 372pp; English.
 XX The invention relates to 592 polynucleotides which have been derived from
 CC a variety of human tissue sources and which encode novel secreted
 CC proteins. The polynucleotides can be used as probes for the
 CC identification and isolation of full length cDNA and genomic DNA. The
 CC polynucleotides and proteins can also be used as nutritional supplements.
 CC The proteins are useful in the treatment of various immune deficiencies
 CC and disorders such as viral infections, bacterial infections, fungal
 CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
 CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
 CC and conditions (e.g. asthma). They are also useful for treating
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
 CC useful for tissue regeneration, for wound healing and in the treatment of
 CC burns, incisions and ulcers. The proteins are also useful for regulating
 CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
 CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention.
 XX
 SQ Sequence 2493 BP; 639 A; 520 C; 800 G; 533 T; 1 other;
 Query Match 0.74; Score 20; DB 24; Length 2493;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2476 ATTCTGGGCGCACCTGGTCC 2495
 DB 564 ATTCTGGGCGCACCTGGTCC 545
 RESULT 43
 ABL16061
 ID ABL16061 standard; cDNA; 2706 BP.
 XX
 AC ABL16061;
 XX
 XX 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 42665.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX

OS Drosophila melanogaster.
 XX
 FN WO200171042-A2.
 EN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) ⁹PE COR¹ NY.
 PI Venter JC, Adams M, Li PMD, Myers EW;
 DR MPI; 2001-656860/75.
 DR P-PSCB; ABB71958.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Claim 1; SEQ ID NO 42665; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB16511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2706 BP; 786 A; 730 C; 727 G; 463 T; 0 other;
 Query Match 0.74; Score 20; DB 23; Length 2706;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1869 ACACGACGAGAAAAAGGAG 1888
 DB 1565 ACACGACGAGAAAAAGGAG 1584
 RESULT 44
 ABL16060/c
 ID ABL16060 standard; cDNA; 4757 BP.
 XX
 AC ABL16060;
 XX
 XX 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 42662.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 FN WO200171042-A2.
 EN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX

AC ABS99353;
 XX 18-DEC-2002 (first entry)
 XX
 XX
 DE Enterococcus faecalis contig sequence #621.
 XX
 XX Computer readable medium; Enterococcus faecalis; microbe; growth;
 KM pathogenicity; vaccine; resistance; Enterococcal infection; commercial;
 KM therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;
 KM biotech technology; antibacterial; modulator of nucleic acid expression;
 KM contig; ds.
 XX
 XX Enterococcus faecalis.
 OS
 XX US2002120116-A1.
 PN
 XX 29-AUG-2002.
 PD
 XX
 XX 04-MAY-1998; 98US-0070927.
 PF
 XX
 XX 04-MAY-1998; 98US-0070927.
 PR
 XX
 XX (KUNS/) KUNSCH C A.
 PA (DILL/) DILLON P J.
 PA (BARA/) BARASH S.
 XX
 PI Kunsch CA, Dillon PJ, Barash S;
 XX
 XX WPI; 2002-750065/81.
 DR
 XX
 XX Computer readable medium having recorded on it a Enterococcus faecalis
 PT nucleic acid sequence useful for detecting diseases related to
 PT Enterococcus infections in animals
 PT
 PS Claim 1; Page -; 119pp; English.
 XX
 XX The present invention relates to a new computer readable medium with an
 CC Enterococcus faecalis nucleic acid sequence. The invention is useful to
 CC diagnose the presence of E.faecalis in a sample or determining the
 CC presence of a specific microbe in a sample. The invention is also useful
 CC for modulating the growth or pathogenicity of E.faecalis, in a vaccine
 CC to confer resistance to Enterococcal infection, for commercial,
 CC therapeutic and industrial purposes, and for fermenting a particular
 CC sugar source or to produce a particular metabolite. The invention is
 CC useful for detecting diseases related to Enterococcus infections in
 CC animals, and for detecting E.faecalis using biotech technology. The
 CC present nucleic acid sequence represents an Enterococcus faecalis contig
 CC DNA sequence of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification but was obtained in electronic format directly
 CC from USPTO at <http://seqdata.uspto.gov>.
 XX
 SQ Sequence 1635 BP; 570 A; 298 C; 308 G; 448 T; 11 other;
 Query Match 0.7%; Score 20; DB 24; Length 1635;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 38 TGGTCTTCTTCAGATTCT 57
 DB 766 TGGTCTTCTTCAGATTCT 785

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX
 XX Drmanac ET, Lie C, Tang YT;
 PI
 XX WPI; 2001-639362/73.
 DR P-PCR; ABG25814.
 CR
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PT
 XX
 PS Claim 1; SEQ ID No 25805; 103pp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantifying a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp://wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1817 BP; 380 A; 492 C; 520 G; 425 T; 0 other;
 Query Match 0.7%; Score 20; DB 23; Length 1817;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2476 ATTCTGGGACCTGCGCC 2495
 DB 287 ATTCTGGGACCTGCGCC 306

RESULT 42
 AAK34875/C
 ID AAK34875 standard; cDNA; 2493 BP.
 XX
 AC AAK34875;
 XX
 XX 08-MAY-2002 (first entry)
 DT
 XX
 DE Human cDNA encoding secreted protein #13.
 XX
 XX Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
 KM viral infection; bacterial infection; fungal infection; diabetes; asthma;

CY 38 TGCTGCTTTTCAGATTCT 57
Db 726 TGCTGCTTTTCAGATTCT 745

RESULT 38

AA553120
ID AA553120 standard; DNA; 1299 BP.

AC AA553120;

DT 13-FEB-2002 (first entry)

DE Enterococcus faecalis DNA for cellular proliferation protein #548.

KM Antisense; ds; prokaryotic cellular proliferation gene;

KW antibiotic; antibacterial; drug design.

OS Enterococcus faecalis.

PN M0200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001MO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Hasselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX P-PSDB; AAU35261.

PS Claim 27; Seq ID No 6757; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

XX prokaryotic cellular proliferation, their use in identifying the

XX genes themselves and the encoded proteins. The prokaryotes used are

XX *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*

XX *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The

XX invention is also useful for the identification of potential new targets

XX for antibiotic development. The antisense nucleic acids can also be used

XX to identify proteins used in proliferation, to express these proteins,

XX and to obtain antibodies capable of binding to the expressed proteins.

XX The proteins can be used to screen compounds in rational drug discovery

XX programmes. The antisense nucleic acid sequence is also useful to screen

XX for homologous nucleic acids which are required for cell proliferation in

XX a wide variety of organisms. The present sequence encodes an

XX essential prokaryotic cellular proliferation protein.

XX Note: The sequence data for this patent did not form part

XX of the printed specification, but was obtained in electronic

XX format directly from WIPO at

XX http://wipo.int/pub/published_pcl_sequences.

XX Sequence 1299 BP; 432 A; 250 C; 268 G; 349 T; 0 other;

XX Query Match 0.7%; Score 20; DB 23; Length 1299;

XX Best Local Similarity 100.0%; Pred. No. 55;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 39 TGCTGCTTTTCAGATTCT 57
Db 726 TGCTGCTTTTCAGATTCT 745

RESULT 39

AA13558
ID AA13558 standard; DNA; 1635 BP.

AC AA13558;

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:621.

KM Enterococcus faecalis; contig; detection; Enterococcal infection;

KW vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN M09850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; 98MO-US08985.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Dillon PJ, Kunsch CA;

XX WPI; 1999-045171/04.

XX New isolated Enterococcus faecalis polynucleotides and polypeptides

XX - used to develop products for the detection of Enterococcus

XX infection.

XX Claim 1; Page 1866-1867; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it

XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

XX AA13558 to AA13919 represent these nucleotide sequences which are

XX primary nucleotide sequences, also known as contigs. The computer-based

XX system can identify fragments of the Enterococcus faecalis genome with

XX commercial importance. The products can be used to detect the presence

XX of Enterococcus faecalis in samples. They can also be used for

XX diagnosing Enterococcal infection in an animal and monitoring

XX progression of disease, and for identifying agents which can be used to

XX modulate the growth or pathogenicity of Enterococcus faecalis, or

XX another related organism, in vivo or in vitro. In particular the

XX polypeptides encoded by the Enterococcus faecalis nucleotide sequences

XX can be used in vaccines to prevent or attenuate an Enterococcal

XX infection.

XX Sequence 1635 BP; 570 A; 298 C; 308 G; 448 T; 11 other;

XX Query Match 0.7%; Score 20; DB 20; Length 1635;

XX Best Local Similarity 100.0%; Pred. No. 56;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX RESULT 40

XX ABS99353

XX ID ABS99353 standard; DNA; 1635 BP.

PR 29-OCT-1999; 99US-0162142.
 Query Match 0.7%; Score 20; DB 21; Length 638;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 746 TTTTCCCAAAAAGATATAA 765
 ID TTTTCCCAAAAAGATATAA 41

RESULT 36
 AAF92514
 ID AAF92514 standard; DNA; 954 BP.
 XX AAF92514;
 XX 16-MAY-2001 (first entry)
 DE Human T2R14 nucleotide sequence SEQ ID NO:27.
 XX Human; rat; mouse; T2R; taste receptor; G-protein coupled receptor;
 KW taste transduction G-protein coupled receptor; identification; tongue;
 KW taste sensory neuron; taste cell; taste modulator; food;
 KW taste signalling pathway; ds.
 XX Homo sapiens.
 OS WO200118050-A2.
 PN 15-MAR-2001.
 PD 08-SEP-2000; 2000WO-US24821.
 PF 10-SEP-1999; 99US-0393634.
 PR 22-FEB-2000; 2000US-0510332.
 XX (REGC) UNIV CALIFORNIA.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Zuker CS, Adler JE, Ryba N, Mueller K, Hoon M;
 PI WPI; 2001-211396/21.
 DR P-PSDB; AAB87744.
 XX Nucleic acids encoding the T2R family of G-protein coupled taste
 PT receptors, useful for identifying taste modulators that can be used in
 PT food and pharmaceutical industries to customize taste, for e.g. to
 PT decrease the bitter taste of food.
 PS Disclosure; Page 171-172; 249pp; English.
 XX AAF92502 to AAF92572 represent nucleic acids which encode taste
 CC transduction G-protein coupled receptors designated T2R proteins.
 CC AAB87731 to AAB87824 represent T2R proteins, and AAB87825 to AAB87830
 CC represent T2R family consensus sequences from the present invention.
 CC The T2R proteins are taste modulators. The nucleic acids are useful as
 CC probes for the identification of taste cells, as the nucleic acids are
 CC specifically expressed in taste cells. They also serve as tools for the
 CC generation of taste topographic maps that elucidate the relationship
 CC between the taste cells of the tongue and taste sensory neurons leading
 CC to taste centres in the brain. The taste modulators are useful for
 CC pharmacological and genetic modulation of taste signalling pathways.
 CC Modulatory compounds comprising T2R proteins can therefore be used in
 CC food and pharmaceutical industries to customise taste, for e.g. to
 CC decrease the bitter taste of food or drugs.

Sequence 954 BP; 249 A; 183 C; 194 G; 328 T; 0 other;

Query Match 0.7%; Score 20; DB 22; Length 954;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TGGTTTGCTGCTGCTCTT 48
 ID TGGTTTGCTGCTGCTCTT 405

RESULT 37 0
 AAS51375
 ID AAS51375 standard; DNA; 1287 BP.
 XX AAS51375;
 XX 13-FEB-2002 (first entry);
 DE Enterococcus faecalis DNA for cellular proliferation protein #152.
 XX Antisense; ds; prokaryotic cellular proliferation gene;
 KW antibiotic; antibacterial; drug design.
 XX Enterococcus faecalis.
 OS WO200170955-A2.
 PN 27-SEP-2001.
 PD 21-MAR-2001; 2001WO-US09180.
 PF 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX (ELITR) ELITRA PHARM INC.
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 PI WPI; 2001-611495/70.
 DR P-PSDB; AAU33516.
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PT Claim 27; Seq ID No 3957; 511pp; English.
 PS The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

Sequence 1287 BP; 422 A; 249 C; 268 G; 348 T; 0 other;

Query Match 0.7%; Score 20; DB 23; Length 1287;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139753.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-01452913.
 PR 27-JUL-1999; 99US-0145918.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149375.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 25-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0153363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 26-SEP-1999; 99US-0156459.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158363.
 PR 13-OCT-1999; 99US-0158293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.

PS Example 4, Page 213, 378pp: English.

XX The invention relates to isolated lung tumour-specific proteins and
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and
 CC their antigen-presenting cells are useful for stimulating and/or
 CC expanding T cells specific for a tumour protein, and for inhibiting
 CC the development of cancer. The invention also relates to a composition
 CC useful for stimulating an immune response, and for treating cancer. The
 CC lung tumour specific oligonucleotide is useful in gene therapy and for
 CC diagnosis, detection and treatment of lung cancer. The present sequence
 CC is a cDNA encoding human lung tumour-specific protein.

XX
 SQ Sequence 548 BP, 180 A; 97 C; 180 G; 91 T; 0 other;

Query Match 0.7%; Score 20; DB 23; Length 548;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2476 ATTCTGGGCGACCTGTCC 2495
 Db 543 ATTCTGGGCGACCTGTCC 524

RESULT 34
 ABL83527
 ID ABL83527 standard; cDNA; 577 BP.

XX
 AC ABL83527;
 DT 17-MAY-2002 (first entry)

XX Human ovarian cancer related cDNA clone SEQ ID NO:6505.
 DE Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
 XX
 KM Homo sapiens.
 OS
 XX WO200192581-A2.
 PN
 XX 06-DEC-2001.
 PD

XX 29-MAY-2001; 2001WO-US17756.
 PF
 XX 26-MAY-2000; 2000US-207484P.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Algate PA, Harlocker SL, Jones R;
 PI
 XX WPI, 2002-122075/16.
 DR

XX Composition for therapy and diagnosis of ovarian cancer comprising
 PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
 PT polypeptide, antibody specific to polypeptide or T cell expressing
 PT polypeptide.
 PT

PS Claim 1; SEQ ID 6505; 489pp; English.

XX The present invention describes a composition (I) comprising: carriers
 CC and immunostimulants; and a polypeptide (III) of a ovarian tumor
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence
 CC (SI) from the 10912 nucleotide sequences as given in ABL77023 to
 CC ABL7934, (III) encoding (II) having a sequence (S2), a T cell
 CC population of (II), or antigen presenting cells that express (II).
 CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridizes to
 CC (SI) can be used for detecting ovarian cancer in a patient's biological
 CC sample preferably serum or ovarian tissue. The method comprises
 CC contacting a biological sample from a patient with (IV), detecting the
 CC amount of polynucleotide hybridizing to (IV) and comparing the amount to
 CC a predetermined cutoff value and thereby detecting ovarian cancer in the
 CC patient, where the amount of polynucleotide hybridizing to (IV) is
 CC detected preferably by polymerase chain reaction (PCR). (I) comprising
 CC (III) and/or (II) is useful for stimulating and/or expanding T cells

CC specific for an ovarian tumour protein comprising contacting T cells
 CC with (III) or (II). (III) is useful in design and preparation of
 CC ribozyme molecules for inhibiting expression of the tumour polypeptides
 CC and proteins in tumour cells; and to isolate a full length gene from a
 CC suitable library e.g., a tumour cDNA library using well known
 CC techniques.

XX
 SQ Sequence 577 BP, 152 A; 121 C; 117 G; 187 T; 0 other;

Query Match 0.7%; Score 20; DB 24; Length 577;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2505 AGAGGCTCCGACGATAC 2524
 Db 345 AGAGGCTCCGACGATAC 364

RESULT 35
 AAC38427
 ID AAC38427 standard; DNA; 639 BP.

XX
 AC AAC38427;
 DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 20930.
 DE Arabidopsis thaliana.
 OS
 XX Arabidopsis thaliana.
 PN
 XX EP1033405-A2.
 PD

XX 06-SEP-2000.
 PF
 XX 25-FEB-2000; 2000EP-0301439.
 PD

XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131445.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134266.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.

```

PF 26-JAN-1999; 99MO-US01642.
XX
PR 22-DEC-1998; 98US-0219245.
PR 28-JAN-1998; 98US-0015022.
PR 28-JAN-1998; 98US-0015029.
PR 18-MAR-1998; 98US-0040828.
PR 18-MAR-1998; 98US-0040831.
PR 23-JUL-1998; 98US-0122191.
PR 23-JUL-1998; 98US-0122192.
XX
PA (CORI-) CORIXA CORP.
PI Fradakis TN, Lodes MJ, Mohamath R, Reed SG;
XX WPI: 1999-479187/40.
DR P-PSDB; AAY29561.
XX
PT Lung tumour specific polynucleotides for inhibiting the development
PT of lung cancer
XX
PS Claim 13; Page 144; 171pp; English.
XX
CC The present invention describes lung tumour specific polynucleotides
CC and tumour antigens. AA207144 to AA207246 and AA208301 to AA208325
CC represent specifically claimed polynucleotides, and AAY29486 to AAY29571
CC represent amino acid sequences from the present invention. The lung
CC tumour specific polynucleotides and polypeptides can be used in
CC pharmaceutical compositions and vaccines to inhibit the development of
CC lung cancer. They can also be used to detect lung cancer in a patient.
CC Probes and antibodies derived from the lung tumour sequences are useful
CC in detection of lung cancer.
XX
SQ Sequence 548 BP; 180 A; 97 C; 180 G; 91 T; 0 other;

Query Match 0.7%; Score 20; DB 20; Length 548;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2476 ATTCCTGGGCCACCTGCTCC 2495
DB 543 ATTCCTGGGCCACCTGCTCC 524

RESULT 32
AAC79168/c
ID AAC79168 standard; cDNA; 548 BP.
XX
AC AAC79168;
XX
DT 05-FEB-2001 (first entry)
XX
DE Human lung tumour-specific cDNA #121.
XX
KW Lung tumour protein; lung cancer; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200060077-A2.
XX
PD 12-OCT-2000.
XX
PE 30-MAR-2000; 2000MO-US08560.
XX
PR 02-APR-1999; 99US-0285323.
PR 09-AUG-1999; 99US-0370838.
PR 30-DEC-1999; 99US-0476235.
PR 03-MAR-2000; 2000US-0518909.
XX
PA (CORI-) CORIXA CORP.
PI Reed SG, Lodes MJ, Mohamath R, Secrist H;
XX WPI: 2000-638466/61.
DR

```

```

DR P-PSDB; AAB44485.
XX
PT Novel lung tumor polypeptides and polynucleotides, useful for
PT detecting, monitoring or treating cancer, especially lung cancer -
XX
PS Claim 1; Page 159; 243pp; English.
XX
CC The present sequence is given in a specification relating to compounds
CC for therapy and diagnosis of lung cancer. Polypeptides comprising at
CC least an immunogenic part of a lung tumour protein are disclosed.
CC The polypeptides are useful for inhibiting the development of cancer,
CC especially lung cancer. Samples of T cells expressing the polypeptides
CC may be used to inhibit the development of cancer. The polypeptides are
CC also useful for detecting and monitoring the progression of cancer,
CC especially lung cancer.
XX
SQ Sequence 548 BP; 180 A; 97 C; 180 G; 91 T; 0 other;

Query Match 0.7%; Score 20; DB 21; Length 548;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2476 ATTCCTGGGCCACCTGCTCC 2495
DB 543 ATTCCTGGGCCACCTGCTCC 524

RESULT 33
AAD23243/c
ID AAD23243 standard; cDNA; 548 BP.
XX
AC AAD23243;
XX
DT 26-FEB-2002 (first entry)
XX
DE Human lung tumour-specific protein SAL-68 cDNA.
XX
KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
XX antisense-therapy; vaccine; immune response; lung cancer; SAL-68; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH CDS 37..547
FH FT /*tag= a
FH FT /product= "Human SAL-68 protein"
FH FT /note= "CDS does not include stop codon"
FH FT /partial
XX
PN WO200172295-A2.
XX
PD 04-OCT-2001.
XX
PE 28-MAR-2001; 2001MO-US09991.
XX
PR 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX
PA (CORI-) CORIXA CORP.
PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indrias CY;
PI Henderson RA, Fling SF, Algate PA, Elliot M, Wammon J, Kalos MD;
XX WPI: 2001-639201/73.
DR P-PSDB; AAE13826.
XX
PT New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer -
XX

```

PS Claim 1; SEQ ID No 5973; 658bp; English.

XX

CC The invention relates to a single exon nucleic acid probe (SENP) (i) for

CC measuring human gene expression in a sample derived from human adult

CC liver, comprising one of 13109 defined nucleotide sequences given in the

CC specification (or complements/ fragments). The probe hybridises at high

CC stringency to a nucleic acid molecule expressed in the human adult

CC liver. (i) may be used for predicting, measuring and displaying gene

CC expression in samples derived from human adult liver. The genes

CC identified may be involved in genetic liver diseases such as cirrhosis,

CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which

CC is associated with coronary heart disease. ABS25011-ABS51005 represent

CC human liver single exon nucleic acid probes of the invention.

CC Note: The sequence information for this patent does not appear in the

CC printed specification but was obtained in electronic format directly

CC from WIPO.int/pub/published_pct_sequences.

XX

SC Sequence 479 BP, 169 A; 82 C; 105 G; 123 T; 0 other:

Query Match 0.7%; Score 20; DB 23; Length 479;

Best Local Similarity 100.0%; Pred. No. 54;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TGGTTTGGTCTGCTCTT 48

DB 386 TGGTTTGGTCTGCTCTT 367

RESULT 30

ABS06054/c

ID ABS06054 standard; DNA: 479 BP.

XX

AC ABS06054;

XX

DT 19-AUG-2002 (first entry)

XX

DE Human genome-derived single exon probe from lung SEQ ID No 6045.

XX

KM Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;

KM chronic obstructive pulmonary disease; interstitial lung disease;

KM familial idiopathic pulmonary fibrosis; neurofibromatosis;

KM tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

KM Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

KM pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;

KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

KM primary ciliary dyskinesia; pulmonary hypertension;

KM hyaline membrane disease.

XX

OS Homo sapiens.

XX

PN WO200186003-A2.

XX

PD 15-NOV-2001.

XX

PF 30-JAN-2001; 2001WO-US00665.

XX

PR 04-FEB-2000; 2000US-180312P.

PR 26-MAY-2000; 2000US-207456P.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632386.

PR 21-SEP-2000; 2000US-234687P.

PR 27-SEP-2000; 2000US-236359P.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SQ, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2002-114183/15.

XX

PT Spatially-addressable set of single exon nucleic acid probes, used to

PT measure gene expression in human lung samples -

XX

PS Claim 1; SEQ ID No 6045; 634bp; English.

XX

CC The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived

CC from human lung comprising single exon nucleic acid probes having one of

CC 12614 nucleic acid sequences mentioned in the specification, or their

CC complements or the 12387 open reading frames derived from the 12614

CC probes. Also included are a microarray comprising the novel set of

CC probes; the novel set of probes which hybridise at high stringency to a

CC nucleic acid expressed in the human lung; measuring gene expression in a

CC sample derived from human lung, comprising (a) contacting the array with

CC a collection of detectably labeled nucleic acids derived from human lung

CC mRNA, and (b) measuring the label detectably bound to each probe of

CC the array; identifying exons in a eukaryotic genome, comprising

CC (a) algorithmically predicting at least one exon from genomic sequences

CC of the eukaryote; and (b) detecting specific hybridisation of detectably

CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,

CC having a fragment identical to the predicted exon, the probe is included

CC in the above mentioned microarray; assigning exons to a single gene,

CC comprising (a) identifying exons from genomic sequence by the method

CC above and (b) measuring the expression of each of the exons in several

CC tissues and/or cell types using hybridisation to a single exon

CC microarray having a probe with the exon, where a common pattern of

CC expression of the exons in the tissues and/or cell types indicates that

CC the exons should be assigned to a single gene; a peptide comprising one

CC of 12011 sequences, mentioned in the specification, or encoded by the

CC probes/open reading frames (ORF). The probes are used for gene

CC expression analysis, and for identifying exons in a gene, particularly

CC using human lung derived mRNA and for the study of lung diseases

CC such as asthma, lung cancer, chronic obstructive pulmonary disease

CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary

CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,

CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary

CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,

CC pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic

CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension

CC and hyaline membrane disease. The present sequence is a single exon

CC probe of the invention.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX

SC Sequence 479 BP, 169 A; 82 C; 105 G; 123 T; 0 other:

Query Match 0.7%; Score 20; DB 24; Length 479;

Best Local Similarity 100.0%; Pred. No. 54;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TGGTTTGGTCTGCTCTT 48

DB 386 TGGTTTGGTCTGCTCTT 367

RESULT 31

AAZ08316/c

ID AAZ08316 standard; CDNA: 548 BP.

XX

AC AAZ08316;

XX

DT 13-OCT-1999 (first entry)

XX

DE Human lung tumour protein SAT-68 5' CDNA sequence.

XX

KM Human; lung tumour protein; therapy; diagnosis; lung cancer; vaccine;

KM immunotherapy; detection; inhibition, ss.

XX

OS Homo sapiens.

XX

PN WO9338973-A2.

XX

PD 05-AUG-1999.

XX

```

DE human bone marrow expressed single exon probe SEQ ID NO: 5859.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KM microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
XX MO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO: 5859; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
XX Sequence 479 BP; 169 A; 82 C; 105 G; 123 T; 0 other;
XX
XX
XX Query Match 0.7%; Score 20; DB 22; Length 479;
XX Best Local Similarity 100.0%; Pred. No. 54;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 29 TGGTTTGGTCTCTCTT 48
XX 386 TGGTTTGGTCTCTCTT 367
XX
XX
XX RESULT 28
XX AA137199/c
XX ID AA137199 standard; DNA; 479 BP.
XX
XX AA137199;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #5885 used to measure gene expression in human placenta sample.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX MO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX
XX

```

```

PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta.
XX
XX Claim 25; SEQ ID No 5885; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 479 BP; 169 A; 82 C; 105 G; 123 T; 0 other;
XX
XX
XX Query Match 0.7%; Score 20; DB 22; Length 479;
XX Best Local Similarity 100.0%; Pred. No. 54;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 29 TGGTTTGGTCTCTCTT 48
XX 386 TGGTTTGGTCTCTCTT 367
XX
XX
XX RESULT 29
XX ABS30983/c
XX ID ABS30983 standard; DNA; 479 BP.
XX
XX ABS30983;
XX
XX 25-FEB-2003 (first entry)
XX
XX Human liver single exon probe, SEQ ID No 5973.
XX
XX Human; single exon nucleic acid probe; liver; cirrhosis;
XX hyperlipoproteinemia; hyperlipidaemia; hypercholesterolaemia;
XX coronary heart disease; ss.
XX
XX Homo sapiens.
XX
XX MO200157273-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00664.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human adult liver.
XX

```

CC with increased tolerance to abiotic stress. The present sequence is that
 CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
 CC in methods of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office.
 XX
 SQ Sequence 393 BP; 133 A; 81 C; 92 G; 87 T; 0 other;

Query Match 0.7%; Score 20; DB 24; Length 393;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 746 TTTCCACCAAAAAGATATAA 765
 Db 33 TTTCCACCAAAAAGATATAA 52

RESULT 25

ABAS7633/C
 ID ABAS7633 standard; DNA: 479 BP.

AC ABAS7633;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #5938.

KM Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human foetal liver -

PS Claim 1; SEQ ID NO 5938; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC foetal liver. The present sequence is a single exon nucleic acid

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 479 BP; 169 A; 82 C; 105 G; 123 T; 0 other;

Query Match 0.7%; Score 20; DB 22; Length 479;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 TGGTTTGGTGGCTTCTT 48
 Db 396 TGGTTTGGTGGCTTCTT 367

RESULT 26

AAK05687/C
 ID AAK05687 standard; DNA: 479 BP.

AC AAK05687;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe SEQ ID NO: 5678.

KM Human; brain expressed exon; gene expression analysis; probe;

KM microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

OS Homo sapiens.

PN WO200157275-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483446/52.

PT Single exon nucleic acid probes for analyzing gene expression in human

PT brains -

PS Example 4; SEQ ID NO: 5678; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC brain. They can be used to measure gene expression in brain cells, samples,

CC which may enable the diagnosis and improved treatment of nervous system

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

CC epilepsy and cancers. The present sequence is one of the probes of the

CC invention.

SQ Sequence 479 BP; 169 A; 82 C; 105 G; 123 T; 0 other;

Query Match 0.7%; Score 20; DB 22; Length 479;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 TGGTTTGGTGGCTTCTT 48
 Db 396 TGGTTTGGTGGCTTCTT 367

RESULT 27

AAK1302/C
 ID AAK1302 standard; DNA: 479 BP.

AC AAK1302;

DT 06-NOV-2001 (first entry)

PR	22-JUL-1999	99US-0145088
PR	22-JUL-1999	99US-0145087
PR	22-JUL-1999	99US-0145089
PR	22-JUL-1999	99US-0145152
PR	23-JUL-1999	99US-0145145
PR	23-JUL-1999	99US-0145148
PR	23-JUL-1999	99US-0145276
PR	26-JUL-1999	99US-0145274
PR	27-JUL-1999	99US-0145513
PR	27-JUL-1999	99US-0145518
PR	27-JUL-1999	99US-0145519
PR	28-JUL-1999	99US-0145551
PR	02-AUG-1999	99US-0146386
PR	02-AUG-1999	99US-0146388
PR	03-AUG-1999	99US-0146389
PR	04-AUG-1999	99US-0147038
PR	04-AUG-1999	99US-0147204
PR	05-AUG-1999	99US-0147192
PR	05-AUG-1999	99US-0147260
PR	06-AUG-1999	99US-0147303
PR	06-AUG-1999	99US-0147416
PR	09-AUG-1999	99US-0147493
PR	09-AUG-1999	99US-0147335
PR	10-AUG-1999	99US-0148171
PR	11-AUG-1999	99US-0148319
PR	12-AUG-1999	99US-0148341
PR	13-AUG-1999	99US-0148365
PR	13-AUG-1999	99US-0148684
PR	16-AUG-1999	99US-0148684
PR	17-AUG-1999	99US-0149175
PR	18-AUG-1999	99US-0149176
PR	20-AUG-1999	99US-0149722
PR	20-AUG-1999	99US-0149723
PR	20-AUG-1999	99US-0149929
PR	23-AUG-1999	99US-0149302
PR	23-AUG-1999	99US-0149330
PR	25-AUG-1999	99US-0150566
PR	26-AUG-1999	99US-0150684
PR	27-AUG-1999	99US-0151065
PR	27-AUG-1999	99US-0151080
PR	27-AUG-1999	99US-0151086
PR	30-AUG-1999	99US-0151303
PR	31-AUG-1999	99US-0151308
PR	01-SEP-1999	99US-0151330
PR	07-SEP-1999	99US-0151263
PR	10-SEP-1999	99US-0153708
PR	13-SEP-1999	99US-0153758
PR	15-SEP-1999	99US-0154018
PR	16-SEP-1999	99US-0154039
PR	22-SEP-1999	99US-0154779
PR	22-SEP-1999	99US-0155133
PR	24-SEP-1999	99US-0155486
PR	28-SEP-1999	99US-0156458
PR	29-SEP-1999	99US-0156596
PR	04-OCT-1999	99US-0157117
PR	05-OCT-1999	99US-0157865
PR	06-OCT-1999	99US-0158029
PR	07-OCT-1999	99US-0158232
PR	08-OCT-1999	99US-0158369
PR	12-OCT-1999	99US-0158369
PR	13-OCT-1999	99US-0159293
PR	13-OCT-1999	99US-0159294
PR	14-OCT-1999	99US-0159329
PR	14-OCT-1999	99US-0159330
PR	14-OCT-1999	99US-0159637
PR	14-OCT-1999	99US-0159638
PR	18-OCT-1999	99US-0159684
PR	21-OCT-1999	99US-0160741

PR	21-Oct-1599	99US-0160767
PR	21-Oct-1599	99US-0160768
PR	21-Oct-1599	99US-0160770
PR	21-Oct-1599	99US-0160814
PR	21-Oct-1599	99US-0160815
PR	22-Oct-1599	99US-0160980
PR	22-Oct-1599	99US-0160981
PR	22-Oct-1599	99US-0160989
PR	22-Oct-1599	99US-0161048
PR	25-Oct-1599	99US-0161405
PR	25-Oct-1599	99US-0161406
PR	25-Oct-1599	99US-0161406
PR	26-Oct-1599	99US-0161159
PR	26-Oct-1599	99US-0161160
PR	26-Oct-1599	99US-0161361
PR	28-Oct-1599	99US-0161240
PR	28-Oct-1599	99US-0161992
PR	28-Oct-1599	99US-0162993
PR	29-Oct-1599	99US-0162142

Query Match	0.7%;	Score 20;	DB 2i;	Length 393;
Best Local Similarity	100.0%;	Pred. No. 54;		
Matches	20;	Conservative	3;	Mismatches 0;
				Indels 0;
				Gaps 0;

Oy	746	TTTCACCCAAAAAAGATAAAA	7655
Db	33	TTTCACCCAAAAAAGATAAAA	52

RESULT 24
ABZ14620
ID ABZ14620 standard; DNA; 393 BP.

DT	21-JAN-2003 (first entry)
XX	
DE	Arabidopsis thaliana stress regulated gene SEQ ID NO 2425.

KW *Arabidopsis thaliana*; plant; gene; stress; transgenic; ds.

OS *Arabidopsis thaliana*.

PN W0200216655-A2.

PD 28-FEB-2002.

PF 24-AUG-2001; 2001WO-US26685.

PR 24-AUG-2000; 2000US-227866P.

PR 22-JUN-2001; 2001US-300111P

PA (Scripps Res Inst.

[illegible]

XX

XX

and producing plants with increased tolerance to these abiotic stressors

XX

XX

cell has been exposed, comprising:

in the plant cell with an array of probes representative of the plant

(b) detecting a profile of expressed polynucleotides in the plant cell

production of transgenic plants, cells and seeds and in producing plants

PT Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 35858; 3071bp + Sequence listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (II) proteins and polynucleotides may be used to prevent
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 2154 BP; 638 A; 443 C; 480 G; 593 T; 0 other;
Query Match 0.8%; Score 22; DB 22; Length 2154;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1178 CTGGAAGAAGAACTGTCA 1199
|||
Db 268 CTGGAAGAAGAACTGTCA 289
RESULT 23
AAC43283
ID AAC43283 standard; DNA; 393 BP.
XX
AC AAC43283;
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 36698.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0128234.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140692.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142355.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144333.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.

```
XX 17-JAN-2001; 2001WO-US01354.
PF
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225287.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
XX

PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 09-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251946.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-C259678.
XX

PA (HUMA-), HUMAN GENOME SCI INC.
XX
PI Rosen CA, Bairash SC, Ruben SM;
XX
DR MPI; 2001-483426/52.
XX
```



```

OS Homo sapiens.
XX
XX Bp1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 33926; 71bp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX
XX Sequence 246 BP; 70 A; 53 C; 71 G; 40 T; 2 other:
XX
XX Query Match 4.6%; Score 131; DB 21; Length 246;
XX Best Local Similarity 100.0%; Pred. No. 1.6e-52;
XX Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Db 2722 GACCCAGGTCTCCCTGGCAAGATGAGACCATGAGAAACCTGGAATCCAGGGACCA 2781
XX |||||
XX 112 GACCCAGGTCTCCCTGGCAAGATGAGACCATGAGAAACCTGGAATCCAGGGACCA 271
XX
XX Qy 2782 GGGCCCCCAGGCGATCTGGCAACCATCACTATGTTTAACTTAATTCGAGAGAGATCCG 2841
XX |||||
XX Db 172 GGGCCCCCAGGCGATCTGGCAACCATCACTATGTTTAACTTAATTCGAGAGAGATCCG 231
XX
XX Qy 2842 TTCAGAAAAGG 2852
XX |||||
XX Db 232 TTCAGAAAAGG 242
XX
XX RESULT 19
XX ABA69546/c
XX ID ABA69546 standard; DNA; 181 BP.
XX
XX AC ABA69546;
XX
XX 31-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #17851.
XX
XX Human: foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX MO2001.57277-A2.
XX
XX 09-AUG-2001.

```

```

XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180112.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (XOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
XX Claim 4; SEQ ID NO 17851; 639bp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 181 BP; 46 A; 45 C; 35 G; 55 T; 0 other:
XX
XX Query Match 2.9%; Score 82; DB 22; Length 181;
XX Best Local Similarity 100.0%; Pred. No. 6.8e-29;
XX Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 2599 AGGCTCTCGAGGAAGAAATGGGAAAGGAGCCAGGCTTGAGTATCTCGAGACCA 2657
XX |||||
XX Db 165 AGGCTCTCGAGGAAGAAATGGGAAAGGAGCCAGGCTTGAGTATCTCGAGACCA 105
XX
XX Qy 2658 AGGTCTCTCTGTCCTCCCAAGT 2679
XX |||||
XX Db 105 AGGTCTCTCTGTCCTCCCAAGT 84
XX
XX RESULT 20
XX ABR71706
XX ID ABR71706 standard; cDNA; 1101 BP.
XX
XX AC ABR71706;
XX
XX 30-JUL-2002 (first entry)
XX
XX Human: dithp polynucleotide #172.
XX
XX Human; dithp; diagnostic and therapeutic polynucleotide; gene; ss; bone;
XX cell proliferative disorder; cancer; tumour; autoimmune disorder; seizure;
XX inflammatory disorder; viral infection; bacterial infection;
XX fungal infection; parasitic infections; developmental disorder; breast;
XX endocrine disorder; metabolic disorder; neurological disorder; cervix;
XX gastrointestinal disorder; transport disorder; gene therapy; kidney;
XX adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;
XX skin; testis; thymus.
XX
XX Homo sapiens.
XX
XX MO200220754-A2.
XX
XX 14-MAR-2002.
XX
XX 29-AUG-2001; 2001WO-US27127.

```

XX PN WO9845435-A2.
 XX PD 15-OCT-1998.
 XX PF 10-APR-1998; 98MO-US06954.
 XX PR 10-APR-1997; 97US-0835913.
 XX PA (GENEY) GENETICS INST INC.
 XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JX, Merberg D;
 XX PI Racine LA, Spaulding V, Treacy M;
 XX DR WPI, 1999-070076/06.
 PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 XX PS Claim 1; Page 180; 633pp; English.
 XX CC This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC therapy. The EST sequences are also stated to be useful for gene
 CC therapy.
 XX SQ Sequence 447 BP; 141 A; 80 C; 101 G; 125 T; 0 other;
 Query Match 8.8%; Score 252; DB 20; Length 447;
 Best Local Similarity 99.4%; Pred. No. 7.6e-111;
 Matches 352; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ATGGCTCACTATTACATTTCTCTGCATGGTTTGGTCTCTTCCAGAAATTCGTG 60
 Db 85 ATGGCTCACTATTACATTTCTCTGCATGGTTTGGTCTCTTCCAGAAATTCGTG 144
 QY 61 TTAGCTGAAGATGGGAAAGTAAAGATCAAGTTGTGTACTGCTCCGACAGATTATGTTTC 120
 Db 145 TTAGCTGAAGATGGGAAAGTAAAGATCAAGTTGTGTACTGCTCCGACAGATTATGTTTC 204
 QY 121 ATCTAGATGGCTCTTATAGTGTGTGGCCAGAAAACCTTGAAATAGTGAAGAAAGTGGCTT 180
 Db 205 ATCTAGATGGCTCTTATAGTGTGTGGCCAGAAAACCTTGAAATAGTGAAGAAAGTGGCTT 264
 QY 181 GTCAATATCACAAAAAAGCTTTGACATAGAGGCCGAAGTTTATTCAGAGTTGAGTGTCAA 240
 Db 265 GTCAATATCACAAAAAAGCTTTGACATAGAGGCCGAAGTTTATTCAGAGTTGAGTGTCAA 324
 QY 241 TATAGTGACTACCTGTGCTGTGGAGATTCCTCTCGAAGCTATGTTTACAGAGAACATTGG 300
 Db 325 TATAGTGACTACCTGTGCTGTGGAGATTCCTCTCGAAGCTATGTTTACAGAGAACATTGG 384
 QY 301 ACGGACAGAGTGAATGCATACCTACTTACTTAGAGGAAACACAAAGACAGGGAG 354
 Db 385 ACGGACAGAGTGAATGCATACCTACTTACTTAGAGGAAACACAAAGACAGGGAG 418

RESULT 17
 AA149752/c
 ID AA149752 standard; DNA; 221 BP.
 XX AC AA149752;

XX DT 17-OCT-2001 (first entry)
 XX DE Probe #18438 used to measure gene expression in human placenta sample.
 XX KW Probe; microarray; human; placenta; antenatal diagnosis;
 XX KW genetic disorder; ss.
 XX OS Homo sapiens.
 XX PN WO200157272-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US00653.
 XX PR 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0612356.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0226359.
 XX PR 04-OCT-2000; 2000GB-0024253.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Bern SG, Hanzel DK, Chen W, Rank DR;
 XX PI WPI, 2001-488897/53.
 XX DR WPI, 2001-488897/53.
 XX PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human placenta
 XX PS Claim 25; SEQ ID NO 18438; 654pp; English.
 XX CC The present invention relates to single exon nucleic acid probes (SENP).
 XX CC CC producing a microarray for predicting, measuring and displaying gene
 XX CC expression in samples derived from human placenta. The probes are useful
 XX CC for antenatal diagnosis of human genetic disorders.
 XX SQ Sequence 221 BP; 62 A; 59 C; 43 G; 57 T; 0 other;
 Query Match 5.8%; Score 166; DB 22; Length 221;
 Best Local Similarity 100.0%; Pred. No. 2.2e-69;
 Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2435 GAATTTGTGATCATTTGCTGTGCTCCACATAGGCTCCCGGATATTCCTGSSCCACCTGGTC 2494
 Db 221 GAATTTGTGATCATTTGCTGTGCTCCACATAGGCTCCCGGATATTCCTGSSCCACCTGGTC 162
 QY 2495 CGATAGGCGCCAGAGGGTCCCGAGAGATTACCTGGTTGCGAGAGAGATGGTGTGCTG 2554
 Db 161 CGATAGGCGCCAGAGGGTCCCGAGAGATTACCTGGTTGCGAGAGAGATGGTGTGCTG 102
 QY 2555 GATTAGTGGGTGTCTCTGAGCTGCAGGTCCAGAGATTAAGG 2690
 Db 101 GATTAGTGGGTGTCTCTGAGCTGCAGGTCCAGAGATTAAGG 56

RESULT 18
 AAC29851
 ID AAC29851 standard; cDNA; 246 BP.
 XX AC AAC29851;
 XX DT 06-OCT-2000 (first entry)
 XX DE Human secreted protein; 5' EST; SEQ ID NO: 33926.
 XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 XX KW gene therapy; chromosome mapping; ss.

QY 781 TCAGAGTGTATATACAGACTCAGACGAATGTTCCAGAGGCTCTTCATCA 840
 |||||
 Db 975 TCAGAGTGTATATATACAGACTCAGACGAATGTTCCAGAGGCTCTTCATCA 1034
 |||||
 QY 841 TATGATTTTGTCTACTCAAGATTTAAAGTCAGAAAATTGGATTATGAGATA 900
 |||||
 Db 1035 TATGATTTTGTCTACTCAAGATTTAAAGTCAGAAAATTGGATTATGAGATA 1094
 |||||
 QY 901 TTAAGTATGATGAGAGCCA 921
 |||||
 Db 1095 TTAAGTATGATGAGAGCCA 1115
 |||||
 RESULT 15
 ABK71703
 ID ABK71703 standard; cDNA; 1414 BP.
 XX
 AC ABK71703;
 XX
 DT 30-JUL-2002 (first entry)
 XX
 DE Human dithp polynucleotide #169.
 XX
 KW Human; dithp; diagnostic and therapeutic polynucleotide; gene; ss; bone;
 KW cell proliferative disorder; cancer; tumour; autoimmune disorder; brain;
 KW inflammatory disorder; viral infection; bacterial infection; seizure;
 KW fungal infection; parasitic infections; developmental disorder; breast;
 KW endocrine disorder; metabolic disorder; neurological disorder; cervix;
 KW gastrointestinal disorder; transport disorder; gene therapy; kidney;
 KW adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;
 KW skin; testis; thymus.
 XX
 OS Homo sapiens.
 EN MO200220754-A2.
 PD 14-MAR-2002.
 XX
 PF 29-AUG-2001; 2001MO-US27127.
 XX
 PR 05-SEP-2000; 2000US-229747P.
 PR 05-SEP-2000; 2000US-229748P.
 PR 05-SEP-2000; 2000US-229749P.
 PR 05-SEP-2000; 2000US-229750P.
 PR 05-SEP-2000; 2000US-229751P.
 PR 05-SEP-2000; 2000US-230583P.
 PR 06-SEP-2000; 2000US-230505P.
 PR 06-SEP-2000; 2000US-230514P.
 PR 06-SEP-2000; 2000US-230515P.
 PR 06-SEP-2000; 2000US-230517P.
 PR 06-SEP-2000; 2000US-230518P.
 PR 06-SEP-2000; 2000US-230519P.
 PR 06-SEP-2000; 2000US-230595P.
 PR 06-SEP-2000; 2000US-230597P.
 PR 06-SEP-2000; 2000US-230598P.
 PR 06-SEP-2000; 2000US-230599P.
 PR 06-SEP-2000; 2000US-230610P.
 PR 06-SEP-2000; 2000US-230655P.
 PR 06-SEP-2000; 2000US-230988P.
 PR 07-SEP-2000; 2000US-230951P.
 PR 07-SEP-2000; 2000US-231163P.
 PR 07-SEP-2000; 2000US-231167P.
 XX
 PA (INCYTE GENOMICS INC.
 XX
 PI Stuart J. Lincoln SE. Altus CM. Dufour GE. Chalup YS. Hillman JI;
 PI Jones AL. Yu JY. Wright RJ. Gierzen D. Liu TF. Yap PS. Dahl CR;
 PI Momiyama MG. Bradley DL. Rohatgi SD. Harris B. Rosberry AM;
 PI Gerstlun BH. Peralta CH. David WP. Panzer SR. Flores V. Daffo A;
 PI Kawata R. Chen AJ. Chang SC. Au AP. Inman RR;
 XX
 DR WFI; 2002-383054/41.
 DR P-PSDB; ABG60112.

XX
 PT An isolated polynucleotide useful in diagnostics and therapeutics -
 XX
 PS Claim 1: Page 486; 686pp; English.
 XX
 CC The invention relates to human diagnostic and therapeutic (dithp)
 CC polynucleotides and their associated polypeptides (dithp polypeptides).
 CC The sequences of the invention are used in the treatment and diagnosis of
 CC cell proliferative disorders (e.g. atherosclerosis, cirrhosis), cancers
 CC (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,
 CC cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or
 CC thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,
 CC psoriasis, osteoporosis), viral infections, bacterial infections, fungal
 CC infections, parasitic infections, developmental disorders (e.g. anaemia,
 CC epilepsy), seizure disorders (e.g. cerebral palsy, spina bilida),
 CC endocrine disorders (e.g. thrombosis, aneurysm), metabolic disorders
 CC (e.g. obesity, diabetes), neurological disorders (e.g. stroke,
 CC amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal
 CC disorders (e.g. ulcerative colitis, lysinuria) and transport disorders
 CC (e.g. myotonic dystrophy, cataracta, peripheral neuropathy). Sequences
 CC ABK71535-ABK71809 represent human dithp polynucleotides of the invention.
 XX
 SC Sequence 1414 BP; 425 A; 267 C; 295 G; 427 T; 0 other;
 XX
 Query Match 13.3%; Score 382; DB 24; Length 1414;
 Best Local Similarity :00.0%; Pred. No. 1.7e-173;
 Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2397 AGCCAGTACCACTCTTACTTCAGAGTGCAGAGATTGAAATTGATGATTCCTGTC 2456
 |||||
 Db 959 AGCCAGTACCACTCTTACTTCAGAGTGCAGAGATTGAAATTGATGATTCCTGTC 1018
 |||||
 QY 2457 CCACATGCGTCCCGGGGTATCTCTGGGCCCACTGGTCCGATAGGCCCAAGGCTCCGAG 2516
 |||||
 Db 1019 CCACATGCGTCCCGGGGTATCTCTGGGCCCACTGGTCCGATAGGCCCAAGGCTCCGAG 1078
 |||||
 QY 2517 AGGATTACCTGTTGGCAGAGAGAGATGGTCTCTGATTAAGTGGGTCTCCCTGGAGC 2576
 |||||
 Db 1079 AGGATTACCTGTTGGCAGAGAGAGATGGTCTCTGATTAAGTGGGTCTCCCTGGAGC 1138
 |||||
 QY 2577 TCAGAGTGTCAAGAGATTAAAGGCTTACCTCAGAGTGCAGAGAAATGGGAAAAAGGAGCCCAAG 2636
 |||||
 Db 1139 TCAGAGTGTCAAGAGATTAAAGGCTTACCTCAGAGTGCAGAGAAATGGGAAAAAGGAGCCCAAG 1198
 |||||
 QY 2637 GTTGGGATCTCTGAGAGACAGGTCTCTGGTCCCGCCAGGTCCAGAGGCGCTCTG 2696
 |||||
 Db 1199 GTTGGGATCTCTGAGAGACAGGTCTCTGGTCCCGCCAGGTCCAGAGGCGCTCTG 1258
 |||||
 QY 2697 AATAAGCAAGAGGCTCTCCAGAGACCCAGGTCTCTGCGCAAGATGAGAGCATGG 2756
 |||||
 Db 1259 AATAAGCAAGAGGCTCTCCAGAGACCCAGGTCTCTGCGCAAGATGAGAGCATGG 1318
 |||||
 QY 2757 AAAACCTGGAATCCAAAGGCAA 2778
 |||||
 Db 1319 AAAACCTGGAATCCAAAGGCAA 1340
 |||||
 RESULT 16
 AA986274
 ID AA986274 standard; cDNA; 447 BP.
 XX
 AC AA986274;
 XX
 DT 27-APR-1999 (first entry)
 XX
 DE EST clone AJ54.
 XX
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 XX
 OS Homo sapiens.

KM acquired immune deficiency syndrome; AIDS; autoimmune disorder;
 KM respiratory disorder; ss.
 XX Homo sapiens.
 OS WO200162927-A2.
 PN 30-AUG-2001.
 XX 21-FEB-2001; 2001WO-US06059.
 PF 24-FEB-2000; 2000US-0184693.
 XX 24-FEB-2000; 2000US-0184697.
 XX 24-FEB-2000; 2000US-0184698.
 XX 24-FEB-2000; 2000US-0184768.
 XX 24-FEB-2000; 2000US-0184769.
 XX 24-FEB-2000; 2000US-0184770.
 XX 24-FEB-2000; 2000US-0184771.
 XX 24-FEB-2000; 2000US-0184772.
 XX 24-FEB-2000; 2000US-0184773.
 XX 24-FEB-2000; 2000US-0184776.
 XX 24-FEB-2000; 2000US-0184777.
 XX 24-FEB-2000; 2000US-0184797.
 XX 24-FEB-2000; 2000US-0184813.
 XX 24-FEB-2000; 2000US-0184837.
 XX 24-FEB-2000; 2000US-0184841.
 XX 24-FEB-2000; 2000US-0185213.
 XX 24-FEB-2000; 2000US-0185216.
 XX 12-MAY-2000; 2000US-0203785.
 XX 15-MAY-2000; 2000US-0204226.
 XX 16-MAY-2000; 2000US-0204525.
 XX 16-MAY-2000; 2000US-0204821.
 XX 16-MAY-2000; 2000US-0204908.
 XX 16-MAY-2000; 2000US-0205232.
 XX 17-MAY-2000; 2000US-0204815.
 XX 17-MAY-2000; 2000US-0204863.
 XX 17-MAY-2000; 2000US-0205221.
 XX 17-MAY-2000; 2000US-0205285.
 XX 17-MAY-2000; 2000US-0205286.
 XX 17-MAY-2000; 2000US-0205287.
 XX 17-MAY-2000; 2000US-0205323.
 XX 17-MAY-2000; 2000US-0205324.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
 P1 Chen A, D'Sa SA, Amshley S, Dahl CR, Dam TC, Daniels SE;
 P1 Dutour GE, Flores V, Fong WT, Greenwalt JB, Hillman UL, Jones AL;
 P1 Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdresher TK, Delfo A;
 P1 Wright RJ, Yap PE, Yu JY, Bradley DJ, Bratcher SR, Chen W;
 P1 Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
 XX WPI: 2001-502867/55.
 DR P-PDB3; AAU19561.
 XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
 PT enzymes, hormones and receptors, useful in diagnostics and therapeutics
 -
 Claim 1; Page 365; 522pp; English.
 CC The invention relates to polynucleotides (I) encoding diagnostic and
 CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
 CC and proteins involved in growth and development and receptors. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate DITHP expression. For example, (I) and
 CC (II) may be used to treat disorders associated with decreased polypeptide
 CC expression by rectifying mutations or deletions in a patient's genome,
 CC that affect the activity of the DITHPs, by expressing inactive proteins
 CC or supplementing the patient's own production of them. (I) and (II)
 CC may be used to treat diseases, for example, cell proliferative disorder,
 CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,

CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
 CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
 CC into a host cell and culturing the cell to express the protein. (I) and
 CC its complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative
 CC therapy. (II) may also be used as antigens in the production of
 CC antibodies against DITHPs and in assays to identify modulators of DITHP
 CC expression and activity. The anti-DITHP antibodies and antagonists may
 CC also be used to down regulate expression and activity. The anti-DITHP
 CC antibodies may also be used as diagnostic agents for detecting the
 CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbent
 CC assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and
 CC therapeutic (DITHP) polynucleotides of the invention.
 CC XX

SQ Sequence 1115 BP; 344 A; 215 C; 259 G; 296 T; 1 other;
 Query Match 26.1%; Score 749; DB 22; Length 1115;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 919; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY	1	ATGGCTCAGTATTATTCATTCTGTCAGTGGTTGGTGGCTCTTCAGAAATTCGTG	60
DB	196	ATGGCTCAGTATTATTCATTCTGTCAGTGGTTGGTGGCTCTTCAGAAATTCGTG	255
QY	61	TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTCGTGCTCCGACAGATTTAGTTTC	120
DB	236	TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTCGTGCTCCGACAGATTTAGTTTC	315
QY	121	ATCTTAGATGGCTCTTATAGTGTGGCCGAGAACTTGAATAGTGAAGTGGCTT	180
DB	316	ATCTTAGATGGCTCTTATAGTGTGGCCGAGAACTTGAATAGTGAAGTGGCTT	375
QY	181	GTCATATACAAAAAACTTTGACATAGGCGCCAAAGTTTATCAAGTTGAGTGTCTCA	240
DB	376	GTCATATACAAAAAACTTTGACATAGGCGCCAAAGTTTATCAAGTTGAGTGTCTCA	435
QY	241	TATAGTACATACCCCTGTCGAGATTCCTCTCGAAGCTATGATAGAGAAATTTG	300
DB	436	TATAGTACATACCCCTGTCGAGATTCCTCTCGAAGCTATGATAGAGAAATTTG	495
QY	301	ACGGCAGCGTGTGAATTCATCTACTTACAGAGAAACACAAAGAGAGAGGCGCATC	360
DB	496	ACGGCAGCGTGTGAATTCATCTACTTACAGAGAAACACAAAGAGAGAGGCGCATC	555
QY	361	CAGTTGGGCTGCGATTCCTTTTGGCAAGCTCCAGATTCCTGACTGATAGACAGTG	420
DB	556	CAGTTGGGCTGCGATTCCTTTTGGCAAGCTCCAGATTCCTGACTGATAGACAGTG	615
QY	421	GTACTTACGATGGCAATTCACAGATGACCTCAAGATGACCTCAAGCAGCAAGAT	480
DB	616	GTACTTACGATGGCAATTCACAGATGACCTCAAGATGACCTCAAGCAGCAAGAT	675
QY	481	AGTAAGATTAACATTTTGTGATATGCTGTGTTGTTCAAGAAAGATGCCAACTTGA	540
DB	676	AGTAAGATTAACATTTTGTGATATGCTGTGTTGTTCAAGAAAGATGCCAACTTGA	735
QY	541	GCTATGGCAAGACCTTCGTACTATGTTTATGTAAGAGATATATGCAATA	600
DB	736	GCTATGGCAAGACCTTCGTACTATGTTTATGTAAGAGATATATGCAATA	795
QY	601	TCCAAATTAAGGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATG	660
DB	796	TCCAAATTAAGGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATG	855
QY	661	ATTCAGAGGAGGCTCGATGATGATGATGATGATGATGATGATGATGATGATG	720
DB	856	ATTCAGAGGAGGCTCGATGATGATGATGATGATGATGATGATGATGATGATG	914
QY	721	AAAAAGTTAAGAAAGATACAGCTTTCACCAAAAAAATTAAGAGATATGAAGTACA	780
DB	915	AAAAAGTTAAGAAAGATACAGCTTTCACCAAAAAAATTAAGAGATATGAAGTACA	974

PR 06-SEP-2000; 2000US-230599P.
 PR 06-SEP-2000; 2000US-230610P.
 PR 06-SEP-2000; 2000US-230865P.
 PR 06-SEP-2000; 2000US-230988P.
 PR 06-SEP-2000; 2000US-230989P.
 PR 07-SEP-2000; 2000US-230951P.
 PR 07-SEP-2000; 2000US-231163P.
 PR 07-SEP-2000; 2000US-231167P.
 PA (INCY-) INCYTE GENOMICS INC.

PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL,
 PI Jones AL, Yu JY, Wright RJ, Gierzen D, Liu TF, Yap PE, Dahn CR,
 PI Moriyma MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM,
 PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Dafio A,
 PI Matwana R, Chen AJ, Chang SC, Au AP, Imman RR;
 XX MPI: 2002-527544/56.
 DR P-PSDB; ABP51293.

PT Novel human disease detection and treatment polypeptide, useful in
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
 PT e.g. AIDS

XX Claim 1; Page 341; 618pp; English.

XX The invention relates to an isolated human disease detection and
 CC treatment (MDT) polypeptide (I) selected from a polypeptide having a
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
 CC specification, a naturally occurring polypeptide comprising a sequence
 CC having at least 90% identity to (I) or a biologically active or
 CC immunogenic fragment of (I). (I) is useful for screening a compound for
 CC effectiveness as an agonist or antagonist for screening a compound that
 CC specifically binds (I) or modulates the activity of (I), and for
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
 CC screening a compound for effectiveness in altering expression of a target
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
 CC detecting MDT in a sample or for assessing toxicity of a test compound,
 CC in a diagnostic test for a condition or a disease associated with the
 CC expression of MDT in a biological sample, for detecting (I) in a sample,
 CC and for purifying (I) from a sample. A composition comprising (I), an
 CC agonist or antagonist is useful for treating a disease or condition
 CC associated with decreased or increased expression of functional MDT.
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders
 CC associated with aberrant expression of MDT, where the disorders are
 CC selected from a cell proliferative disorder such as arteriosclerosis,
 CC cirrhosis, hepatitis, psoriasis, and cancer and an
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
 CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised
 CC animals or transgenic animals to model human diseases, in somatic or
 CC germ-line gene therapy, to generate a transcript image of a tissue or cell
 CC type, for detecting differences in the chromosomal location due to cell
 CC translocation or inversion among normal, carrier or affected individuals
 CC and as hybridisation probes for mapping naturally occurring genomic
 CC sequences.

XX Sequence 1169 BP; 364 A; 224 C; 267 G; 314 T; 0 other;

XX Query Match 32.2%; Score 923; DB 24; Length 1169;
 XX Best Local Similarity 99.9%; Pred. No. 0;
 XX Matches 973; Conservative 0; Mismatches 1; Indels 0; Gaps 3;

QY 1 ATGGCTCACTAATTACATTCTCGATGCTTTGGTGGTCTCTTCAGAAATCTGTG 60
 Db 196 ATGGCTCACTAATTACATTCTCTCGATGCTTTGGTGGTCTCTTCAGAAATCTGTG 255
 QY 61 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTCGACTGCTCCGACAGATTAGTTTC 120
 Db 256 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTCGACTGCTCCGACAGATTAGTTTC 315

QY 121 ATCTTAGATGGCTCTTAAGTGTGGCCCAAGAACTTGAATAGTAAAAAGTGGCTT 180
 Db 316 ATCTTAGATGGCTCTTAAGTGTGGCCCAAGAACTTGAATAGTAAAAAGTGGCTT 375
 QY 181 GTCATATACAAAAGAACTTGCATATAGGGCCGAAGTTATTCAGTTGGAGTGGTCAA 240
 Db 370 GTCATATACAAAAGAACTTGCATATAGGGCCGAAGTTATTCAGTTGGAGTGGTCAA 435
 QY 241 TATAGTACTACCTCTGCTGAGATTCCTCTCGAAGCTATGATTCAGAGACATTTG 300
 Db 436 TATAGTACTACCTCTGCTGAGATTCCTCTCGAAGCTATGATTCAGAGACATTTG 495
 QY 301 ACGGACAGTGAATTCATCTCTACTTATGAGAGAAACAAAGACAGGAGGCAATC 360
 Db 496 ACGGACAGTGAATTCATCTCTACTTATGAGAGAAACAAAGACAGGAGGCAATC 555
 QY 361 CAGTTGCGCTGATACCTTTTGGCAAGTCTCGATTTCTGACTAAGATAGAGTG 420
 Db 556 CAGTTGCGCTGATACCTTTTGGCAAGTCTCGATTTCTGACTAAGATAGAGTG 615
 QY 421 GTAATTGAGATGCAATTCGCAAGATGAGTCAAGATGACGCTCAAGCAGAGAGAT 480
 Db 616 GTAATTGAGATGCAATTCGCAAGATGAGTCAAGATGACGCTCAAGCAGAGAGAT 675
 QY 481 AGTAGAGTACATCATTTGCTATGCTGTGTTGCAAGAGATGCGAAGCTTAA 540
 Db 676 AGTAGAGTACATCATTTGCTATGCTGTGTTGCAAGAGATGCGAAGCTTAA 735
 QY 541 GCTATTGCAAGAGCTTCTGCTACTTATGCTTTATGTTATGAGACATATTCGCAAT 600
 Db 736 GCTATTGCAAGAGCTTCTGCTACTTATGCTTTATGTTATGAGACATATTCGCAAT 795
 QY 601 TCCAAATTAAGGAAGTGAAGAGCAAACTTGTGAAGATTCGTCTGCCACACAGA 660
 Db 796 TCCAAATTAAGGAAGTGAAGAGCAAACTTGTGAAGATTCGTCTGCCACACAGA 855
 QY 661 ATTCAGTGGACGCTGTATGAAGAGGATTTGATTTCTTTAGTTAGATGAAT 720
 Db 856 ATTCAGTGGACGCTGTATGAAGAGGATTTGATTTCTTTAGTTAGATGAAT 915
 QY 721 AAAAAGTTAAGAAAGATACGCTTTTACCCAAAAGATTAAGATATGAATACA 780
 Db 916 AAAAAGTTAAGAAAGATACGCTTTTACCCAAAAGATTAAGATATGAATACA 975
 QY 780 TCAAAGTATTATTCAGACACACAGCAATGTTTCCGAAAGTCTTCCATCA 840
 Db 976 TCAAAGTATTATTCAGACACACAGCAATGTTTCCGAAAGTCTTCCATCA 1035
 QY 841 TATGATTTGTGCTTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGAGAAAT 900
 Db 1036 TATGATTTGTGCTTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGAGAAAT 1095
 QY 901 TTACTATGATGAGAGGCAAAATGAGTACCTTAAAGTGTGAGCAAAATCTTA 960
 Db 1096 TTACTATGATGAGAGGCAAAATGAGTACCTTAAAGTGTGAGCAAAATCTTA 1155
 QY 961 TTATTTACACAAAC 974
 Db 1156 TTATTTACACAAAC 1169

RESULT 14
 AAS31132
 ID AAS31132 standard; CEVA; 1115 BP.
 XX AAS31132;
 XX 04-DEC-2001 (first entry)
 XX Human diagnostic and therapeutic polynucleotide (DITHP) #147.
 XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
 XX cell proliferative disorder; Crohn's disease; lymphoma; leukemia;

PD 09-AUG-2001.
 XX 05-FEB-2001; 2001WO-US04098.
 XX 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX (HXYE-) HXYE INC.
 PA Tang YT, Liu C, Dimaac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZM;
 P1 Xue AJ, Yang Y, Wehrman T, Goodrich R;
 XX WPI: 2001-476283/51.
 DR E-PSDB; AAM79896.
 DR Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 PT Claim 1; Page 4838; 6221pp; English.
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haemopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEO ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 XX Sequence 2209 BP; 674 A; 413 C; 544 G; 577 T; 1 other;
 SQ
 Query Match 33.7%; Score 966; DB 22; Length 2209;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 527 GATGGCTTGATGGGGCCCCCAGGCTCTAAGGGGCAACCTGGAGATCCAGTCTCAGGGA 586
 QY CCCCAGGTTGGATGGGAAGCCGGAAGAGGTTTCAACAAATTAATTCAGCAAGT 2379
 DB CCCCAGGTTGGATGGGAAGCCGGAAGAGGTTTCAACAAATTAATTCAGCAAGT 646
 QY TGCAAGATTAATTAAGAGCCGAGTACGATCTTACTTCAAGTGGAAATTAAGT 2439
 DB TGCAAGATTAATTAAGAGCCGAGTACGATCTTACTTCAAGTGGAAATTAAGT 706
 QY TGTGATATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2495
 DB TGTGATATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 766
 QY GGGCCAGAGGATCCGAGAGATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2559
 DB GGGCCAGAGGATCCGAGAGATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 826
 QY GTGGGTGTCCCTGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGG 2619
 DB GTGGGTGTCCCTGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGG 886
 QY GAAAAAGAGAGCCAGAGGCTTGGGTCCTGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGG 2675
 DB GAAAAAGAGAGCCAGAGGCTTGGGTCCTGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGG 946
 QY 2680 CCAAGAGGCTTCTGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGG 2739
 DB CCAAGAGGCTTCTGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGG 1006
 QY 2740 AAGATGAGAGCAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGG 2799
 DB AAGATGAGAGCAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGG 1066
 QY 2800 GACCCATCACTATGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2859
 DB GACCCATCACTATGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1126
 QY 2860 TATTAG 2865
 DB TATTAG 1132
 DE
 XX Human polynucleotide SEQ ID NO 4177.
 XX
 XX Human: neotrophic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW Leukemia; ss.
 XX
 XX Homo sapiens.
 OS
 XX W020015312-A1.
 PN
 XX 26-JUL-2001.
 PD
 XX 26-DEC-2000; 2000WO-US34263.
 PF
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620325.

```

PR 07-SEP-2000; 2000US-231167P.
PA (INCYTE GENOMICS INC.
XX
XX Jackson S., Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
XX Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TP, Yap PE, Dahl CR;
XX Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
XX Gerstein EH, Peralta CH, David WH, Panzer SR, Flores V, Daffo A;
XX Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;
XX MPI: 2002-527544/56.
XX P-PSDB: ABP51423.
XX
XX Novel human disease detection and treatment polypeptide, useful in
XX diagnosis, prevention or treatment of cell proliferative disorders e.g.
XX arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
XX e.g. AIDS
XX
XX Claim 1, Page 409-410; 618pp; English.
XX
XX The invention relates to an isolated human disease detection and
XX treatment (MDOT) polypeptide (I) selected from a polypeptide having a
XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the
XX specification, a naturally occurring polypeptide comprising a sequence
XX having at least 90% identity to (I) or a biologically active or
XX immunogenic fragment of (I). (I) is useful for screening a compound for
XX effectiveness as an agonist or antagonist, for screening a compound that
XX specifically binds (I) or modulates the activity of (I), and for
XX preparing a polyclonal or monoclonal antibody by hybridoma technology.
XX Nucleic acids (II) (ABQ2449-ABQ2700) encoding (I) are useful for
XX screening a compound for effectiveness in altering expression of a target
XX polynucleotide comprising. Oligonucleotides and antibodies are useful for
XX detecting MDOT in a sample or for assessing toxicity of a test compound,
XX in a diagnostic test for a condition or a disease associated with the
XX expression of MDOT in a biological sample, for detecting (I) in a sample,
XX and for purifying (I) from a sample. A composition comprising (I), an
XX agonist or antagonist is useful for treating a disease or condition
XX associated with decreased or increased expression of functional MDOT.
XX (I) or (II) are useful for diagnosing, treating or preventing disorders
XX associated with aberrant expression of MDOT, where the disorders are
XX selected from a cell proliferative disorder such as arteriosclerosis,
XX cirrhosis, hepatitis, psoriasis, and cancer and an
XX autoimmune/inflammatory disorder such as AIDS, Addison's disease,
XX allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
XX rheumatoid arthritis. (II) are useful for creating knockin humanised
XX animals or transgenic animals to model human diseases, in somatic or
XX germline gene therapy, to generate a transcript image of a tissue or cell
XX type, for detecting differences in the chromosomal location due to
XX translocation or inversion among normal, carrier or affected individuals
XX and as hybridisation probes for mapping naturally occurring genomic
XX sequences.
XX
XX Sequence 1184 BP; 368 A; 227 C; 272 G; 317 T; 0 other:
XX
XX Query Match 34.0%; Score 974; DB 24; Length 1184;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ATGGCTCACTATATACATTTCTCGCAGTGGTTTGCGTCTTCCAGAAATTCGTG 60
XX |
XX Db 211 ATGGCTCACTATATATACATTTCTCGCAGTGGTTTGCGTCTTCCAGAAATTCGTG 276
XX |
XX Qy 61 TTAGCTGAAGTGGGGAAGTAAAGATCAAGTTGTCTACTGCTCCGACAGATTTGATTTTC 120
XX |
XX Db 271 TTAGCTGAAGTGGGGAAGTAAAGATCAAGTTGTCTACTGCTCCGACAGATTTGATTTTC 336
XX |
XX Qy 121 ATCTAGATGGCTCTTATAGTGTGGCCGAGAAAATTTGAAATATGAGAAAATGGCTT 180
XX |
XX Db 331 ATCTAGATGGCTCTTATAGTGTGGCCGAGAAAATTTGAAATATGAGAAAATGGCTT 390
XX |
XX Qy 181 GTCAATATCACAAAAAATCTTGACATAGGGCCGAGATTTATTCAGTTGAGTGTCTCAA 240
XX |
XX Db 391 GTCAATATCACAAAAAATCTTGACATAGGGCCGAGATTTATTCAGTTGAGTGTCTCAA 450

```

```

Qy 241 TATAGTACTACCCCTGTGCTGAGATTCCTCTCGGAAAGCTATGATTCAGAGAAACATTTG 300
Db 451 TATAGTACTACCCCTGTGCTGAGATTCCTCTCGGAAAGCTATGATTCAGAGAAACATTTG 510
Qy 301 ACGGAGAGATGGAAATGCTATCTTACTTATGAGAGAAACCAAAAGACAGGAGGCGATC 360
Db 511 ACGGAGAGATGGAAATGCTATCTTACTTATGAGAGAAACCAAAAGACAGGAGGCGATC 570
Qy 361 CAGTTTGGCTGATTAACCTTTTGCCAAAGCTCCAGATTTCTGACTTAAGATAGCAGTG 420
Db 571 CAGTTTGGCTGATTAACCTTTTGCCAAAGCTCCAGATTTCTGACTTAAGATAGCAGTG 630
Qy 421 GTACTTACGATGGCAATTCGCAAGATGAGTCAAGATGCTCAAGCAGCAGAT 480
Db 631 GTACTTACGATGGCAATTCGCAAGATGAGTCAAGATGCTCAAGCAGCAGAT 690
Qy 481 AGTAGATTAACATTTATTTGCTATGTTGGTTCCAGAAACAGAAAGATGCCAACTTNGA 540
Db 691 AGTAGATTAACATTTATTTGCTATGTTGGTTCCAGAAACAGAAAGATGCCAACTTNGA 750
Qy 541 GCTATTGGCAACAAAGCTTCTCTACTTATGTTTATGTAAGACTATTTTGCATA 600
Db 751 GCTATTGGCAACAAAGCTTCTCTACTTATGTTTATGTAAGACTATTTTGCATA 810
Qy 601 TCCAAATTAAGGAAAGTATGATGACAGCAAACTTGTGAAAGATGTCCTGCCAAGCAG 660
Db 811 TCCAAATTAAGGAAAGTATGATGACAGCAAACTTGTGAAAGATGTCCTGCCAAGCAG 870
Qy 661 ATTCAGATGGAGGCTGCTGATGAAAGGGGATTTGATATCTTTTATGATGATGTAAT 720
Db 871 ATTCAGATGGAGGCTGCTGATGAAAGGGGATTTGATATCTTTTATGATGATGTAAT 930
Qy 721 AAAAAGTTAAGAAAGATACAGTTTACCAAAAAAGATTAAGATGAAAGTAA 780
Db 931 AAAAAGTTAAGAAAGATACAGTTTACCAAAAAAGATTAAGATGAAAGTAA 990
Qy 781 TCAAAAGTTAATATGCAATCCAAAGCAATGTTTCCGAAAGGCTTCCGCAATCA 840
Db 991 TCAAAAGTTAATATGCAATCCAAAGCAATGTTTCCGAAAGGCTTCCGCAATCA 1050
Qy 841 TATGATATTTGCTCTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGAGATA 900
Db 1051 TATGATATTTGCTCTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGAGATA 1110
Qy 901 TTAATCTTATGATGAAAGCCCAAAATAGCAATTTACCTTAAATGGTGGCAAAATCTTA 960
Db 1111 TTAATCTTATGATGAAAGCCCAAAATAGCAATTTACCTTAAATGGTGGCAAAATCTTA 1170
Qy 961 TTAATTACACAC 974
Db 1171 TTAATTACACAC 1184

```

RESULT 11
 AAK53029
 ID AAK53029 standard; cDNA; 2209 BP.
 XX AAK53029;
 XX 06-NOV-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 2598.
 XX
 XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukemia;
 KM nervous system disorder; arthritis; inflammation; ss.
 CS Homo sapiens.
 XX
 XX WO200157190-A2.

DR WPI: 2001-442253/47.
 DR P-PSDB: AAM39246.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 605; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (A157798-A161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, Amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 XX
 SO Sequence 2230 BP; 679 A; 409 C; 550 G; 592 T; 0 other;

Query Match 37.1%; Score 1062; DB 22; Length 2230;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1062; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1804 GGAATCCAGGATTTCTGTGAAACCGAGATTATGGGCGAAAGAGAAATTGGGCT 1863
 DB 91 GGAATCCAGGATTTCTGTGAAACCGAGATTATGGGCGAAAGAGAAATTGGGCT 150
 QY 1864 CCAGGACAGCAAGGAAAAAGAGAGCCCAAGGATGCTGTTTAAATGGAGCAATGTC 1923
 DB 151 CCAGGACAGCAAGGAAAAAGAGAGCCCAAGGATGCTGTTTAAATGGAGCAATGTC 210
 QY 1924 TCACCGGCGCAGCTGTGAAACCGGAGTCTAAGGAGCAAAAGGTAACCTGGAATTCAA 1983
 DB 211 TCACCGGCGCAGCTGTGAAACCGGAGTCTAAGGAGCAAAAGGTAACCTGGAATTCAA 270
 QY 1984 GGAATGCTGTGGGCTTCTGTGCTCAAGGAGCAACAGAGCAACGGCTTCCAGAGAA 2043
 DB 271 GGAATGCTGTGGGCTTCTGTGCTCAAGGAGCAACAGAGCAACGGCTTCCAGAGAA 330
 QY 2044 CCAGGATCATGGGTTTACCGGGATTCAGGAAAAAGGGGACAAAGAAATCAAGGT 2103
 DB 331 CCAGGATCATGGGTTTACCGGGATTCAGGAAAAAGGGGACAAAGAAATCAAGGT 390
 QY 2104 GAAAAAGGTATTCAGGGTCAAAAAGGAAAAATGGAAGACAGGAAATTCAGGGCAACAG 2163
 DB 391 GAAAAAGGTATTCAGGGTCAAAAAGGAAAAATGGAAGACAGGAAATTCAGGGCAACAG 450
 QY 2164 GGAATTCAGAGCCATCATGTGTCAAAAAGAGAGAGAGGTGAAAAGGAGAACTGGTGT 2223
 DB 451 GGAATTCAGAGCCATCATGTGTCAAAAAGAGAGAGGTGAAAAGGAGAACTGGTGT 510
 QY 2224 CGAGGTGCATTTGATCAAAAAGGAAATCGGGTGAATGGCTTGAATGGGCGCCGAGGT 2283
 DB 511 CGAGGTGCATTTGATCAAAAAGGAAATCGGGTGAATGGCTTGAATGGGCGCCGAGGT 570
 QY 2284 CCTAAGGGCAACCTGGGGATCAAGTCTCAGGAGACCCCAAGTGTGATGGAGAGCC 2343
 DB 571 CCTAAGGGCAACCTGGGGATCAAGTCTCAGGAGACCCCAAGTGTGATGGAGAGCC 630
 QY 2344 GGAAGAGAGTTTTCAGAACTTTATTCAGCAAGTTGGACAGATGTAATAGAGCCAG 2403
 DB 631 GGAAGAGAGTTTTCAGAACTTTATTCAGCAAGTTGGACAGATGTAATAGAGCCAG 690
 QY 2404 CTACAGTCTTACTTCAGAGTGAAGATTAAGAAATGTGATCATTTGCTGTCCAACT 2463

DB 691 CTACAGTCTTACTTCAGAGTGAAGAAATTAGAAATTTGATCATTTGCTGTCCAACT 750
 QY 2464 GGCTCCCGGGTATTTCTTGGCCACTGATCCGATAGGCCCAAGAGGTCCTCCAGAGATTA 2523
 DB 751 GGCTCCCGGGTATTTCTTGGCCACTGATCCGATAGGCCCAAGAGGTCCTCCAGAGATTA 810
 QY 2524 CTTGTTTGCAGAGAGATGATGTTCTTCTGATTAATGAGGTTGCTCCGAGAGTCAGGT 2583
 DB 811 CTTGTTTGCAGAGAGATGATGTTCTTCTGATTAATGAGGTTGCTCCGAGAGTCAGGT 870
 QY 2584 GTCAGAGATTAAGAGGCTTACAGAGAGAAATGAGGAAAAAGAGAGCCAGAGTTGGG 2643
 DB 871 GTCAGAGATTAAGAGGCTTACAGAGAGAAATGAGGAAAAAGAGAGCCAGAGTTGGG 930
 QY 2644 TATCTTGAGAAACAGGTCCTCTGTGCTCCAGAGGTCCTCCAGAGTTGATTAAC 2703
 DB 931 TATCTTGAGAAACAGGTCCTCTGTGCTCCAGAGGTCCTCCAGAGTTGATTAAC 990
 QY 2704 AAGAGAGTCTCTCAGAGAGCCAGGTCCTCTGCGCAAGATGAGAGCAGTGAAGAACT 2763
 DB 991 AAGAGAGTCTCTCAGAGAGCCAGGTCCTCTGCGCAAGATGAGAGCAGTGAAGAACT 1050
 QY 2764 GGAATGCAAGGCAACCGAGGCCCCCAGGATCTGCGACCATCACTATGTTTACTGA 2823
 DB 1051 GGAATGCAAGGCAACCGAGGCCCCCAGGATCTGCGACCATCACTATGTTTACTGA 1110
 QY 2824 ATTGCCAAGAGATCGGTTTCAGAAAGGCAACATTTAG 2865
 DB 1111 ATTGCCAAGAGATCGGTTTCAGAAAGGCAACATTTAG 1152
 RESULT 10
 ID ABQ72640
 ABQ72640 standard; CDNA; 1184 BP.
 XX
 AC ABQ72640;
 XX
 DT 03-SEP-2002 (first entry)
 XX
 DE Human MDR encoding cDNA SEQ ID NO 192.
 XX
 KW Human; MDR; disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antileukosclerotic;
 KW hepatocellular carcinoma; antiproliferative; cytostatic; anti-HIV;
 KW antiallergic; antineoplastic; antitumor; antitumor; antitumor;
 KW neuroprotective; antineoplastic; antitumor; antitumor; antitumor;
 OS Homo sapiens.
 XX
 PN MO200240715-A2.
 XX
 PD 23-MAY-2002.
 XX
 PF 06-SEP-2001; 2001WO-US27628.
 XX
 PR 06-SEP-2000; 2000US-230505P.
 PR 06-SEP-2000; 2000US-230514P.
 PR 06-SEP-2000; 2000US-230515P.
 PR 06-SEP-2000; 2000US-230517P.
 PR 06-SEP-2000; 2000US-230518P.
 PR 06-SEP-2000; 2000US-230519P.
 PR 06-SEP-2000; 2000US-230520P.
 PR 06-SEP-2000; 2000US-230521P.
 PR 06-SEP-2000; 2000US-230522P.
 PR 06-SEP-2000; 2000US-230523P.
 PR 06-SEP-2000; 2000US-230524P.
 PR 06-SEP-2000; 2000US-230525P.
 PR 06-SEP-2000; 2000US-230526P.
 PR 06-SEP-2000; 2000US-230527P.
 PR 06-SEP-2000; 2000US-230528P.
 PR 06-SEP-2000; 2000US-230529P.
 PR 06-SEP-2000; 2000US-230530P.
 PR 06-SEP-2000; 2000US-230531P.
 PR 06-SEP-2000; 2000US-230532P.
 PR 06-SEP-2000; 2000US-230533P.
 PR 06-SEP-2000; 2000US-230534P.
 PR 06-SEP-2000; 2000US-230535P.
 PR 06-SEP-2000; 2000US-230536P.
 PR 06-SEP-2000; 2000US-230537P.
 PR 06-SEP-2000; 2000US-230538P.
 PR 06-SEP-2000; 2000US-230539P.
 PR 06-SEP-2000; 2000US-230540P.
 PR 06-SEP-2000; 2000US-230541P.
 PR 06-SEP-2000; 2000US-230542P.
 PR 06-SEP-2000; 2000US-230543P.
 PR 06-SEP-2000; 2000US-230544P.
 PR 06-SEP-2000; 2000US-230545P.
 PR 06-SEP-2000; 2000US-230546P.
 PR 06-SEP-2000; 2000US-230547P.
 PR 06-SEP-2000; 2000US-230548P.
 PR 06-SEP-2000; 2000US-230549P.
 PR 06-SEP-2000; 2000US-230550P.
 PR 06-SEP-2000; 2000US-230551P.
 PR 06-SEP-2000; 2000US-230552P.
 PR 06-SEP-2000; 2000US-230553P.
 PR 06-SEP-2000; 2000US-230554P.
 PR 06-SEP-2000; 2000US-230555P.
 PR 06-SEP-2000; 2000US-230556P.
 PR 06-SEP-2000; 2000US-230557P.
 PR 06-SEP-2000; 2000US-230558P.
 PR 06-SEP-2000; 2000US-230559P.
 PR 06-SEP-2000; 2000US-230560P.
 PR 06-SEP-2000; 2000US-230561P.
 PR 06-SEP-2000; 2000US-230562P.
 PR 06-SEP-2000; 2000US-230563P.
 PR 06-SEP-2000; 2000US-230564P.
 PR 06-SEP-2000; 2000US-230565P.
 PR 06-SEP-2000; 2000US-230566P.
 PR 06-SEP-2000; 2000US-230567P.
 PR 06-SEP-2000; 2000US-230568P.
 PR 06-SEP-2000; 2000US-230569P.
 PR 06-SEP-2000; 2000US-230570P.
 PR 06-SEP-2000; 2000US-230571P.
 PR 06-SEP-2000; 2000US-230572P.
 PR 06-SEP-2000; 2000US-230573P.
 PR 06-SEP-2000; 2000US-230574P.
 PR 06-SEP-2000; 2000US-230575P.
 PR 06-SEP-2000; 2000US-230576P.
 PR 06-SEP-2000; 2000US-230577P.
 PR 06-SEP-2000; 2000US-230578P.
 PR 06-SEP-2000; 2000US-230579P.
 PR 06-SEP-2000; 2000US-230580P.
 PR 06-SEP-2000; 2000US-230581P.
 PR 06-SEP-2000; 2000US-230582P.
 PR 06-SEP-2000; 2000US-230583P.
 PR 06-SEP-2000; 2000US-230584P.
 PR 06-SEP-2000; 2000US-230585P.
 PR 06-SEP-2000; 2000US-230586P.
 PR 06-SEP-2000; 2000US-230587P.
 PR 06-SEP-2000; 2000US-230588P.
 PR 06-SEP-2000; 2000US-230589P.
 PR 06-SEP-2000; 2000US-230590P.
 PR 06-SEP-2000; 2000US-230591P.
 PR 06-SEP-2000; 2000US-230592P.
 PR 06-SEP-2000; 2000US-230593P.
 PR 06-SEP-2000; 2000US-230594P.
 PR 06-SEP-2000; 2000US-230595P.
 PR 06-SEP-2000; 2000US-230596P.
 PR 06-SEP-2000; 2000US-230597P.
 PR 06-SEP-2000; 2000US-230598P.
 PR 06-SEP-2000; 2000US-230599P.
 PR 06-SEP-2000; 2000US-230600P.
 PR 06-SEP-2000; 2000US-230601P.
 PR 06-SEP-2000; 2000US-230602P.
 PR 06-SEP-2000; 2000US-230603P.
 PR 06-SEP-2000; 2000US-230604P.
 PR 06-SEP-2000; 2000US-230605P.
 PR 06-SEP-2000; 2000US-230606P.
 PR 06-SEP-2000; 2000US-230607P.
 PR 06-SEP-2000; 2000US-230608P.
 PR 06-SEP-2000; 2000US-230609P.
 PR 06-SEP-2000; 2000US-230610P.
 PR 06-SEP-2000; 2000US-230611P.
 PR 06-SEP-2000; 2000US-230612P.
 PR 06-SEP-2000; 2000US-230613P.
 PR 06-SEP-2000; 2000US-230614P.
 PR 06-SEP-2000; 2000US-230615P.
 PR 06-SEP-2000; 2000US-230616P.
 PR 06-SEP-2000; 2000US-230617P.
 PR 06-SEP-2000; 2000US-230618P.
 PR 06-SEP-2000; 2000US-230619P.
 PR 06-SEP-2000; 2000US-230620P.
 PR 06-SEP-2000; 2000US-230621P.
 PR 06-SEP-2000; 2000US-230622P.
 PR 06-SEP-2000; 2000US-230623P.
 PR 06-SEP-2000; 2000US-230624P.
 PR 06-SEP-2000; 2000US-230625P.
 PR 06-SEP-2000; 2000US-230626P.
 PR 06-SEP-2000; 2000US-230627P.
 PR 06-SEP-2000; 2000US-230628P.
 PR 06-SEP-2000; 2000US-230629P.
 PR 06-SEP-2000; 2000US-230630P.
 PR 06-SEP-2000; 2000US-230631P.
 PR 06-SEP-2000; 2000US-230632P.
 PR 06-SEP-2000; 2000US-230633P.
 PR 06-SEP-2000; 2000US-230634P.
 PR 06-SEP-2000; 2000US-230635P.
 PR 06-SEP-2000; 2000US-230636P.
 PR 06-SEP-2000; 2000US-230637P.
 PR 06-SEP-2000; 2000US-230638P.
 PR 06-SEP-2000; 2000US-230639P.
 PR 06-SEP-2000; 2000US-230640P.
 PR 06-SEP-2000; 2000US-230641P.
 PR 06-SEP-2000; 2000US-230642P.
 PR 06-SEP-2000; 2000US-230643P.
 PR 06-SEP-2000; 2000US-230644P.
 PR 06-SEP-2000; 2000US-230645P.
 PR 06-SEP-2000; 2000US-230646P.
 PR 06-SEP-2000; 2000US-230647P.
 PR 06-SEP-2000; 2000US-230648P.
 PR 06-SEP-2000; 2000US-230649P.
 PR 06-SEP-2000; 2000US-230650P.
 PR 06-SEP-2000; 2000US-230651P.
 PR 06-SEP-2000; 2000US-230652P.
 PR 06-SEP-2000; 2000US-230653P.
 PR 06-SEP-2000; 2000US-230654P.
 PR 06-SEP-2000; 2000US-230655P.
 PR 06-SEP-2000; 2000US-230656P.
 PR 06-SEP-2000; 2000US-230657P.
 PR 06-SEP-2000; 2000US-230658P.
 PR 06-SEP-2000; 2000US-230659P.
 PR 06-SEP-2000; 2000US-230660P.
 PR 06-SEP-2000; 2000US-230661P.
 PR 06-SEP-2000; 2000US-230662P.
 PR 06-SEP-2000; 2000US-230663P.
 PR 06-SEP-2000; 2000US-230664P.
 PR 06-SEP-2000; 2000US-230665P.
 PR 06-SEP-2000; 2000US-230666P.
 PR 06-SEP-2000; 2000US-230667P.
 PR 06-SEP-2000; 2000US-230668P.
 PR 06-SEP-2000; 2000US-230669P.
 PR 06-SEP-2000; 2000US-230670P.
 PR 06-SEP-2000; 2000US-230671P.
 PR 06-SEP-2000; 2000US-230672P.
 PR 06-SEP-2000; 2000US-230673P.
 PR 06-SEP-2000; 2000US-230674P.
 PR 06-SEP-2000; 2000US-230675P.
 PR 06-SEP-2000; 2000US-230676P.
 PR 06-SEP-2000; 2000US-230677P.
 PR 06-SEP-2000; 2000US-230678P.
 PR 06-SEP-2000; 2000US-230679P.
 PR 06-SEP-2000; 2000US-230680P.
 PR 06-SEP-2000; 2000US-230681P.
 PR 06-SEP-2000; 2000US-230682P.
 PR 06-SEP-2000; 2000US-230683P.
 PR 06-SEP-2000; 2000US-230684P.
 PR 06-SEP-2000; 2000US-230685P.
 PR 06-SEP-2000; 2000US-230686P.
 PR 06-SEP-2000; 2000US-230687P.
 PR 06-SEP-2000; 2000US-230688P.
 PR 06-SEP-2000; 2000US-230689P.
 PR 06-SEP-2000; 2000US-230690P.
 PR 06-SEP-2000; 2000US-230691P.
 PR 06-SEP-2000; 2000US-230692P.
 PR 06-SEP-2000; 2000US-230693P.
 PR 06-SEP-2000; 2000US-230694P.
 PR 06-SEP-2000; 2000US-230695P.
 PR 06-SEP-2000; 2000US-230696P.
 PR 06-SEP-2000; 2000US-230697P.
 PR 06-SEP-2000; 2000US-230698P.
 PR 06-SEP-2000; 2000US-230699P.
 PR 06-SEP-2000; 2000US-230700P.
 PR 06-SEP-2000; 2000US-230701P.
 PR 06-SEP-2000; 2000US-230702P.
 PR 06-SEP-2000; 2000US-230703P.
 PR 06-SEP-2000; 2000US-230704P.
 PR 06-SEP-2000; 2000US-230705P.
 PR 06-SEP-2000; 2000US-230706P.
 PR 06-SEP-2000; 2000US-230707P.
 PR 06-SEP-2000; 2000US-230708P.
 PR 06-SEP-2000; 2000US-230709P.
 PR 06-SEP-2000; 2000US-230710P.
 PR 06-SEP-2000; 2000US-230711P.
 PR 06-SEP-2000; 2000US-230712P.
 PR 06-SEP-2000; 2000US-230713P.
 PR 06-SEP-2000; 2000US-230714P.
 PR 06-SEP-2000; 2000US-230715P.
 PR 06-SEP-2000; 2000US-230716P.
 PR 06-SEP-2000; 2000US-230717P.
 PR 06-SEP-2000; 2000US-230718P.
 PR 06-SEP-2000; 2000US-230719P.
 PR 06-SEP-2000; 2000US-230720P.
 PR 06-SEP-2000; 2000US-230721P.
 PR 06-SEP-2000; 2000US-230722P.
 PR 06-SEP-2000; 2000US-230723P.
 PR 06-SEP-2000; 2000US-230724P.
 PR 06-SEP-2000; 2000US-230725P.
 PR 06-SEP-2000; 2000US-230726P.
 PR 06-SEP-2000; 2000US-230727P.
 PR 06-SEP-2000; 2000US-230728P.
 PR 06-SEP-2000; 2000US-230729P.
 PR 06-SEP-2000; 2000US-230730P.
 PR 06-SEP-2000; 2000US-230731P.
 PR 06-SEP-2000; 2000US-230732P.
 PR 06-SEP-2000; 2000US-230733P.
 PR 06-SEP-2000; 2000US-230734P.
 PR 06-SEP-2000; 2000US-230735P.
 PR 06-SEP-2000; 2000US-230736P.
 PR 06-SEP-2000; 2000US-230737P.
 PR 06-SEP-2000; 2000US-230738P.
 PR 06-SEP-2000; 2000US-230739P.
 PR 06-SEP-2000; 2000US-230740P.
 PR 06-SEP-2000; 2000US-230741P.
 PR 06-SEP-2000; 2000US-230742P.
 PR 06-SEP-2000; 2000US-230743P.
 PR 06-SEP-2000; 2000US-230744P.
 PR 06-SEP-2000; 2000US-230745P.
 PR 06-SEP-2000; 2000US-230746P.
 PR 06-SEP-2000; 2000US-230747P.
 PR 06-SEP-2000; 2000US-230748P.
 PR 06-SEP-2000; 2000US-230749P.
 PR 06-SEP-2000; 2000US-230750P.
 PR 06-SEP-2000; 2000US-230751P.
 PR 06-SEP-2000; 2000US-230752P.
 PR 06-SEP-2000; 2000US-230753P.
 PR 06-SEP-2000; 2000US-230754P.
 PR 06-SEP-2000; 2000US-230755P.
 PR 06-SEP-2000; 2000US-230756P.
 PR 06-SEP-2000; 2000US-230757P.
 PR 06-SEP-2000; 2000US-230758P.
 PR 06-SEP-2000; 2000US-230759P.
 PR 06-SEP-2000; 2000US-230760P.
 PR 06-SEP-2000; 2000US-230761P.
 PR 06-SEP-2000; 2000US-230762P.
 PR 06-SEP-2000; 2000US-230763P.
 PR 06-SEP-2000; 2000US-230764P.
 PR 06-SEP-2000; 2000US-230765P.
 PR 06-SEP-2000; 2000US-230766P.
 PR 06-SEP-2000; 2000US-230767P.
 PR 06-SEP-2000; 2000US-230768P.
 PR 06-SEP-2000; 2000US-230769P.
 PR 06-SEP-2000; 2000US-230770P.
 PR 06-SEP-2000; 2000US-230771P.
 PR 06-SEP-2000; 2000US-230772P.
 PR 06-SEP-2000; 2000US-230773P.
 PR 06-SEP-2000; 2000US-230774P.
 PR 06-SEP-2000; 2000US-230775P.
 PR 06-SEP-2000; 2000US-230776P.
 PR 06-SEP-2000; 2000US-230777P.
 PR 06-SEP-2000; 2000US-230778P.
 PR 06-SEP-2000; 2000US-230779P.
 PR 06-SEP-2000; 2000US-230780P.
 PR 06-SEP-2000; 2000US-230781P.
 PR 06-SEP-2000; 2000US-230782P.
 PR 06-SEP-2000; 2000US-230783P.
 PR 06-SEP-2000; 2000US-230784P.
 PR 06-SEP-2000; 2000US-230785P.
 PR 06-SEP-2000; 2000US-230786P.
 PR 06-SEP-2000; 2000US-230787P.
 PR 06-SEP-2000; 2000US-230788P.
 PR 06-SEP-2000; 2000US-230789P.
 PR 06-SEP-2000; 2000US-230790P.
 PR 06-SEP-2000; 2000US-230791P.
 PR 06-SEP-2000; 2000US-230792P.
 PR 06-SEP-2000; 2000US-230793P.
 PR 06-SEP-2000; 2000US-230794P.
 PR 06-SEP-2000; 2000US-230795P.
 PR 06-SEP-2000; 2000US-230796P.
 PR 06-SEP-2000; 2000US-230797P.
 PR 06-SEP-2000; 2000US-230798P.
 PR 06-SEP-2000; 2000US-230799P.
 PR 06-SEP-2000; 2000US-230800P.
 PR 06-SEP-2000; 2000US-230801P.
 PR 06-SEP-2000; 2000US-230802P.
 PR 06-SEP-2000; 2000US-230803P.
 PR 06-SEP-2000; 2000US-230804P.
 PR 06-SEP-2000; 2000US-230805P.
 PR 06-SEP-2000; 2000US-230806P.
 PR 06-SEP-2000; 2000US-230807P.
 PR 06-SEP-2000; 2000US-230808P.
 PR 06-SEP-2000; 2000US-230809P.
 PR 06-SEP-2000; 2000US-230810P.
 PR 06-SEP-2000; 2000US-230811P.
 PR 06-SEP-2000; 2000US-230812P.
 PR 06-SEP-2000; 2000US-230813P.
 PR 06-SEP-2000; 2000US-230814P.
 PR 06-SEP-2000; 2000US-230815P.
 PR 06-SEP-2000; 2000US-230816P.
 PR 06-SEP-2000; 2000US-230817P.
 PR 06-SEP-2000; 2000US-230818P.
 PR 06-SEP-2000; 2000US-230819P.
 PR 06-SEP-2000; 2000US-230820P.
 PR 06-SEP-2000; 2000US-230821P.
 PR 06-SEP-2000; 2000US-230822P.
 PR 06-SEP-2000; 2000US-230823P.
 PR 06-SEP

PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZM;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX MPI: 2001-476283/51.
DR P-PSDB: AAM78912.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 1: Page 2132-2134; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK51435) and the
CC encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, hematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibit activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 2230 BP; 679 A; 409 C; 550 G; 592 T; 0 other:
SQ
XX
XX Query Match 37.1%; Score 1062; DB 22; Length 2230;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1062; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2404 CTACAGCTCTACTTCAGAGTGAAGAGATAGAAATTTGATCATTTGCCGTGCCACAT 2463
DB 631 CTACAGCTCTACTTCAGAGTGAAGAGATAGAAATTTGATCATTTGCCGTGCCACAT 750
QY 2464 GGCCTCCCGGGGTAATCTTGGGGCACTGGTCCGATAGAGCCAGAGAGGATCCAGAGATTA 2523
DB 752 GGCCTCCCGGGGTAATCTTGGGGCACTGGTCCGATAGAGCCAGAGAGGATCCAGAGATTA 810
QY 2524⁶ CCTGGTTGGCAGAGAGATGGTGTCTCGATATGATGAGTGTCCCTGGAGTCAGAGT 2583
DB 811 CCTGGTTGGCAGAGAGATGGTGTCTCGATATGATGAGTGTCCCTGGAGTCAGAGT 870
QY 2584 GTCAGAGATTTAAAGGCTTACAGAGAGAAATGGGAGAAAGAGCCAAAGGCTTGGG 2643
DB 871 GTCAGAGATTTAAAGGCTTACAGAGAGAAATGGGAGAAAGAGCCAAAGGCTTGGG 930
QY 2644 TATCTGGAGAGCAAGTCTCTGGTCCCGAGGTCAGAGAGGCTCTCTGGATTAAC 2703
DB 931 TATCTGGAGAGCAAGTCTCTGGTCCCGAGGTCAGAGAGGCTCTCTGGATTAAC 990
QY 2704 AAAGAGGCTCTCCAGAGAGCCAGAGTCTCTGGCMAAGATGAGACCATGAGAAACCT 2763
DB 991 AAAGAGGCTCTCCAGAGAGCCAGAGTCTCTGGCMAAGATGAGACCATGAGAAACCT 1050
QY 2764 GGAATCCAGAGGCAAGGAGCCAGAGCATGAGACCATGATATATTTAGTGTGA 2823
DB 1051 GGAATCCAGAGGCAAGGAGCCAGAGCATGAGACCATGATATATTTAGTGTGA 1110
QY 2824 ATTGCCAGAGAGATCCGTTTCAAGAAAGACCAACTATTAG 2865
DB 1111 ATTGCCAGAGAGATCCGTTTCAAGAAAGACCAACTATTAG 1152

RESULT 9
AA158402
ID AA158402 standard; cDNA; 2230 BP.
XX
XX AA158402;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 605.
XX
XX Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; hereditary;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thymolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
XX Homo sapiens.
XX
XX W0206153112-A1.
XX
XX 26-OCT-2001.
XX
XX 26-DEC-2003; 2000US-0534263.
XX
XX 21-JAN-2003; 2000US-0488725.
XX 25-APR-2003; 2000US-0552317.
XX 09-JUL-2000; 2000US-0596642.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HISE-) HISEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao Q, Zhou P, Goodrich R, Drmanac RT;

CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAK80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.

XX Sequence 1691 BP; 532 A; 326 C; 386 G; 447 T; 0 other;

Query Match 42.0%; Score 1203; DB 22; Length 1691;
 Best Local Similarity 99.8%; Pred. No. 0;

Matches 1423; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 ATGGCTACATATTAATTAATTTCTGTGATGGTTTGGTGGCTTCTTCAGAAATCTGTG 60
 Db 115 ATGGCTACATATTAATTAATTTCTGTGATGGTTTGGTGGCTTCTTCAGAAATCTGTG 174
 QY 61 TTACCTGAAGATGGGGAGTAAGTCAAGTTTGTCTGCTCCGACAGATTTAGTTTC 120
 Db 175 TTACCTGAAGATGGGGAGTAAGTCAAGTTTGTCTGCTCCGACAGATTTAGTTTC 234
 QY 121 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTGAATAATGAAAGAGGCTT 180
 Db 235 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTGAATAATGAAAGAGGCTT 234
 QY 181 GTCAATATCAAAAAAATTGACATAGGCGCAAGTTTATCAAGTTGAGTGTTCAA 240
 Db 295 GTCAATATCAAAAAAATTGACATAGGCGCAAGTTTATCAAGTTGAGTGTTCAA 354
 QY 241 TATAGTACTACCTGTGCTGGAGATTCCTCTGGAGATGATGATTCAGAGAACTTTG 300
 Db 355 TATAGTACTACCTGTGCTGGAGATTCCTCTGGAGATGATGATTCAGAGAACTTTG 414
 QY 301 ACGGACAGATGAAATCCATATCTTACTTAGAGAGAAACAACAAGACAGGAAGGCCATC 360
 Db 415 ACGGACAGATGAAATCCATATCTTACTTAGAGAGAAACAACAAGACAGGAAGGCCATC 474
 QY 361 CAGTTGGCTCGATTCACCTTTTGGCAAGTCCCTCAGATTTCTGACTAAGATAGCAGT 420
 Db 475 CAGTTGGCTCGATTCACCTTTTGGCAAGTCCCTCAGATTTCTGACTAAGATAGCAGT 534
 QY 421 GTACTTAGAGATGCAATCCCAAGATGATGATGATGATGATGATGATGATGATGAT 480
 Db 535 GTACTTAGAGATGCAATCCCAAGATGATGATGATGATGATGATGATGATGATGAT 594
 QY 481 AGTAAGATTAACATTTTGTCTATTGTGTGTTGTTGAGAAACAGAAATGCCAATTA 540
 Db 595 AGTAAGATTAACATTTTGTCTATTGTGTGTTGTTGAGAAACAGAAATGCCAATTA 654
 QY 541 GCTATTGCCAAGACCTGCTGCTACTTATGTTGTTTATGAGAAAGCTATATTGGAATA 600
 Db 655 GCTATTGCCAAGACCTGCTGCTACTTATGTTGTTTATGAGAAAGCTATATTGGAATA 714
 QY 601 TCCAAAATAAGGAAGTATGATGAACAGAACTTTGTGAAGATCTGCTGCCACAGCA 660
 Db 715 TCCAAAATAAGGAAGTATGATGAACAGAACTTTGTGAAGATCTGCTGCCACAGCA 774
 QY 661 ATTCCAGTGGAGCTGCTGATGAAGGGGATTTGATTTCTTTAGGTTTATGATGAAT 720
 Db 775 ATTCCAGTGGAGCTGCTGATGAAGGGGATTTGATTTCTTTAGGTTTATGATGAAT 834
 QY 721 AAAAAGTTAAGAAAGATACAGCTTACCAAAAAAGATGAAGATGAAGTAA 780
 Db 835 AAAAAGTTAAGAAAGATACAGCTTACCAAAAAAGATGAAGATGAAGTAA 854
 QY 781 TCAAAAGTTGATTTATGAAAGCTCAAGCAATGTTTCCCAAGAGCTTCTCCATCA 840
 Db 895 TCAAAAGTTGATTTATGAAAGCTCAAGCAATGTTTCCCAAGAGCTTCTCCATCA 954
 QY 841 TATGATTTGTGTACTCAAAAGTTTAAAGTCAAGAAATTTGGATTTATGAGAAATA 900

Db 955 TATGATTTGTGCTTACTCAAAAGTTTAAAGTCAAGAAATTTGGGATTTATGAGAAATA 1014
 QY 901 TTAACATATTGATGGAGGCCCAAAATAGCAAGTTACCTTAAATGGTGTGCAAAATCTTA 960
 Db 1015 TTAACATATTGATG-AAGCCAAATAGCAAGTTACCTTAAATGGTGTGCAAAATCTTA 1073
 QY 961 TTAATTACAAACACAGGTTAATTAATGCTGCACAAATGTTTACTTGTCAACCTTCA 1020
 Db 1074 TTAATTACAAACACAGGTTAATTAATGCTGCACAAATGTTTACTTGTCAACCTTCA 1133
 QY 1021 GTTAGACCTGTTTATGATGAAGGCTGGACCAAAATGCTCTTATGATGAAGCAAGAT 1080
 Db 1134 GTTAGACCTGTTTATGATGAAGGCTGGACCAAAATGCTCTTATGATGAAGCAAGAT 1193
 QY 1081 GTGACTTTGATATGATGATGACCAAAATGGAAGGCTTACATCCAGTTTAAAGG 1140
 Db 1194 GTGACTTTGATATGATGATGACCAAAATGGAAGGCTTACATCCAGTTTAAAGG 1253
 QY 1141 ATCTTGATCAATGAGCAAAACCAAAATTTGAAAAATATCTGAAAAAGAAAGAACTGTT 1200
 Db 1254 ATCTTGATCAATGAGCAAAACCAAAATTTGAAAAATATCTGAAAAAGAAAGAACTGTT 1313
 QY 1201 TTTGATGTCAAAAAGTTGCGAATCTACTGTGACCCAGAAAGCAACACCGGAGACAGCA 1260
 Db 1314 TTTGATGTCAAAAAGTTGCGAATCTACTGTGACCCAGAAAGCAACACCGGAGACAGCA 1373
 QY 1261 TGTGAGATTTCTGATTTTGGCTTAAATGCTCCAGTATGATGATCACTCCAGCTGCC 1320
 Db 1374 TGTGAGATTTCTGATTTTGGCTTAAATGCTCCAGTATGATGATGATCACTCCAGCTGCC 1433
 QY 1321 TGTATTTGCTCCGGGGAAAAACAGGACTTCAAGGCTGAAAGGCTGAGCTGGCT 1380
 Db 1434 TGTATTTGCTCCGGGGAAAAACAGGACTTCAAGGCTGAAAGGCTGAGCTGGCT 1453
 QY 1381 GGGAACTCTGCTACCTCTGACAACTTGATGAATGATGATGATGATGATGATGATGAT 1426
 Db 1494 GGGAACTCTGCTACCTCTGACAACTTGATGAATGATGATGATGATGATGATGATGAT 1539

RESULT 8
 ID AAK52045
 XX AAK52045 standard; cDNA; 2230 BP.
 AC AAK52045;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 590.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO20157:90-42.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WC-0504098.
 XX
 PR 03-FEB-2003; 2000US-0496914.
 PR 27-APR-2003; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR -5-SEP-2000; 2000US-0653561.
 PR 20-OCT-2000; 2000US-0693125.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HSE-); HVSSEQ INC.
 XX

QY 61 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTGTACTGTCGAGACAGATTAGTTTC 120
 DB 62 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTGTACTGTCGAGACAGATTAGTTTC 121
 QY 121 ATCTTAGATGCTCTTATAGTGTGGCCAGAAAACTTTGAATAGTGAAGAAAGTGGCTT 180
 DB 122 ATCTTAGATGCTCTTATAGTGTGGCCAGAAAACTTTGAATAGTGAAGAAAGTGGCTT 181
 QY 181 GTCAATATCACAAAACTTTGACACTAGGCGCGAAGTTTATCAAGTTGAGTGGTTCAA 240
 DB 182 GTCAATATCACAAAACTTTGACACTAGGCGCGAAGTTTATCAAGTTGAGTGGTTCAA 241
 QY 241 TATAGTGACTACCTGTGTGTGAGATTCCTCTCGAAGCTATGATTGAGAGAAACATTTG 300
 DB 242 TATAGTGACTACCTGTGTGTGAGATTCCTCTCGAAGCTATGATTGAGAGAAACATTTG 301
 QY 301 ACGGACAGAGTGGAAATCCATACCTTCTTAAAGAGAAACAAAGACAGGAAAGCCATC 360
 DB 302 ACGGACAGAGTGGAAATCCATACCTTCTTAAAGAGAAACAAAGACAGGAAAGCCATC 361
 QY 361 CAGTTTGGCTGATTAACCTTTTTCGCAAGTCTCAGAGTTTCTGACATAAGATAGCAGTG 420
 DB 362 CAGTTTGGCTGATTAACCTTTTTCGCAAGTCTCAGAGTTTCTGACATAAGATAGCAGTG 421
 QY 421 GTACTTACCGATGGCAAAATCCCAAGATGAGTCAAGATGAGTCAAGACAGCAAGAT 480
 DB 422 GTACTTACCGATGGCAAAATCCCAAGATGAGTCAAGATGAGTCAAGACAGCAAGAT 481
 QY 481 AGTAGATTAAGATTAATTTGTGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
 DB 482 AGTAGATTAAGATTAATTTGTGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 541
 QY 541 GGTATTTGCCAAGACCTTCGTCTAATTATGTTTATGTTGAAGACTATATTGCAATA 600
 DB 542 GGTATTTGCCAAGACCTTCGTCTAATTATGTTTATGTTGAAGACTATATTGCAATA 601
 QY 601 TCCAAAATAAGGGAAGTGTATGAAGCAAACTTTGTGAAGAAATGTCTGTCCCAACGCA 660
 DB 602 TCCAAAATAAGGGAAGTGTATGAAGCAAACTTTGTGAAGAAATGTCTGTCCCAACGCA 661
 QY 661 ATTCAGTGGCAGCTGATGATGAAGGGGATTTGATTTCTTTTAGTTAGATGATTAAT 720
 DB 662 ATTCAGTGGCAGCTGATGATGAAGGGGATTTGATTTCTTTTAGTTAGATGATTAAT 721
 QY 721 AAAAAGTTAAGAAAAAATACAGCTTTTACCAAAAAAGATTAAGATGAGTACCA 780
 DB 722 AAAAAGTTAAGAAAAAATACAGCTTTTACCAAAAAAGATTAAGATGAGTACCA 781
 QY 781 TCAAAAGTTGATTTATCAGAACTCAAGCAAGATGTTTCCAGAAAGTCTTCCCACTCA 840
 DB 782 TCAAAAGTTGATTTATCAGAACTCAAGCAAGATGTTTCCAGAAAGTCTTCCCACTCA 841
 QY 841 TATGTAATTTGTGTCTACTCAAAAGATTAAAGTCAGAAAAATTTGGGATTTTGGAGATA 900
 DB 842 TATGTAATTTGTGTCTACTCAAAAGATTAAAGTCAGAAAAATTTGGGATTTTGGAGATA 901
 QY 901 TTAACCTATGATGAGAAAGCCCAATAGCAGTTTAAATGTTGTGAGCAAAATCTTA 960
 DB 902 TTAACCTATGATGAGAAAGCCCAATAGCAGTTTAAATGTTGTGAGCAAAATCTTA 961
 QY 961 TTAATTTAACAACCAAGCGTAATTAATGAGTCAACAAGTGTACTTGTGTAACCTTCAA 1020
 DB 962 TTAATTTAACAACCAAGCGTAATTAATGAGTCAACAAGTGTACTTGTGTAACCTTCAA 1021
 QY 1021 GTTAAGACCTTGTGATGAGAGCTGGACCAAAATTTGCTCTTAGTAACAAGAACAGAT 1080
 DB 1022 GTTAAGACCTTGTGATGAGAGCTGGACCAAAATTTGCTCTTAGTAACAAGAACAGAT 1081
 QY 1081 GTTACTTTGTATATTGATGACCAAAATTTGAAAAAAGCCCTTACATCCAGTTTAGGG 1140
 DB 1082 GTTACTTTGTATATTGATGACCAAAATTTGAAAAAAGCCCTTACATCCAGTTTAGGG 1141

QY 1141 ATCTTAGATGCTCTTATAGTGTGGCCAGAAAACTTTGAATAGTGAAGAAAGTGGCTT 1200
 DB 1141 ATCTTAGATGCTCTTATAGTGTGGCCAGAAAACTTTGAATAGTGAAGAAAGTGGCTT 1201
 QY 1201 TTGTATGCTCCAAAAAGTTGGCAATCTACTGTAGACCCAGAAACAAACCGGAGACAGCA 1260
 DB 1201 TTGTATGCTCCAAAAAGTTGGCAATCTACTGTAGACCCAGAAACAAACCGGAGACAGCA 1261
 QY 1261 TGTAGAGTTCCTGGATTTTGGCTTATAGTCCAGAGTGTAGTGGTCAACTCCAGCTCCC 1320
 DB 1261 TGTAGAGTTCCTGGATTTTGGCTTATAGTCCAGAGTGTAGTGGTCAACTCCAGCTCCC 1321
 QY 1321 TGTATTTGCTCTCCGGAAGAAACAGAGACTTCAAGGCCCAAGAGTGAACCTGACTGCT 1380
 DB 1321 TGTATTTGCTCTCCGGAAGAAACAGAGACTTCAAGGCCCAAGAGTGAACCTGACTGCT 1381
 QY 1381 GGGAAACCTGCTACCTCTGACAACTGTGCAAGATGTAAGCTTG 1426
 DB 1381 GGGAAACCTGCTACCTCTGACAACTGTGCAAGATGTAAGCTTG 1427

RESULT 7

AAK52265 standard; cDNA; 1691 BP.

AAK52265;

06-NOV-2001 (first entry)

Human polynucleotide SEQ ID NO 810.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 human; cytokine; cell proliferation; stem cell growth factor; haematopoiesis;
 tissue growth factor; immunomodulatory; cancer; leukemia;
 nervous system disorder; arthritis; inflammation; se.

Homo sapiens.

WO20015790-A2.

03-NOV-2001.

05-FEB-2001; 2001WO-USC4098.

03-FEB-2000; 2000US-0496914.

27-APR-2000; 2000US-0560875.

20-JUN-2000; 2000US-058075.

19-JUL-2000; 2000US-0620325.

01-SEP-2000; 2000US-0649361.

15-SEP-2000; 2000US-0653561.

20-OCT-2000; 2000US-0693325.

30-NOV-2000; 2000US-0728422.

(HISE-) HISEQ INC.

Tang YT, Liu C, Dymnac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

Zhao QJ, Wang D, Wang J, Zhang Y, Ren F, Chen R, Wang ZW;

Xue AJ, Yang Y, Wejhtman T, Goodrich R;

WPI: 2001-476283/51.

P-PSDB: AAM79132.

Nucleic acids encoding polypeptides with cytokine-like activities,

useful in diagnosis and gene therapy -

Claim 1: Page 27:10-27:11; 6221pp; English.

The invention relates to polynucleotides (AAK51456-AAK51435) and the
 encoded polypeptides (AAM78323-AAK80302) that exhibit activity inducing
 cytokine, cell proliferation or cell differentiation or which may induce
 production of other cytokines in other cell populations. The
 polynucleotides and polypeptides are useful in gene therapy, vaccines or
 peptide therapy. The polypeptides have various cytokine-like activities,

```

Db      1381 GGGAACTCTGCTACCTGGACAACTGGTCAAGATGTAAGCTTG 1426
302 ACCGACAGATGGATCCATCTACTTAGAGGAAACAAAGCAGGGAGGCCATC 361
361 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCCAGATTCTGACTAAGATCAGTG 420
362 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCCAGATTCTGACTAAGATCAGTG 421
421 GTACTTACGATGGCAAAATCCCAAGATGACGTCAAGATGACGTCAAGCAGCAGAGAT 480
422 GTACTTACGATGGCAAAATCCCAAGATGACGTCAAGATGACGTCAAGCAGCAGAGAT 481
481 AGTAAATTAACATTAATTTGCTATTGCTGTTGCTGCAAAACAGAAATCCCAACTTGA 540
482 AGTAAATTAACATTAATTTGCTATTGCTGTTGCTGCAAAACAGAAATCCCAACTTGA 541
541 GCTATTGCCAAACAGCTTCTGCTACTTATGTTTATGTTGGAAGACTATATTGCAATA 600
542 GCTATTGCCAAACAGCTTCTGCTACTTATGTTTATGTTGGAAGACTATATTGCAATA 601
601 TCCAAATTAAGGAAAGTATGATGAAGCAAACTTTGGAAGATCTGCTGTCACAGCA 660
602 TCCAAATTAAGGAAAGTATGATGAAGCAAACTTTGGAAGATCTGCTGTCACAGCA 661
661 ATTCAGTGGAGCTCGTATGAAAGGGGATTTGATATTCTTTAGCTTAGATGTAAT 720
662 ATTCAGTGGAGCTCGTATGAAAGGGGATTTGATATTCTTTAGCTTAGATGTAAT 721
721 AAAAAGGTTAAGAAAGATACAGCTTTCCACAAAAGATTAAGATTAAGTAACTA 780
722 AAAAAGGTTAAGAAAGATACAGCTTTCCACAAAAGATTAAGATTAAGTAACTA 781
781 TCAAAAGTTGATTATGAGAACTCACAAGCAATGTTTCCACAGAGTCTTCTCTCATCA 840
782 TCAAAAGTTGATTATGAGAACTCACAAGCAATGTTTCCACAGAGTCTTCTCTCATCA 841
841 TATGTAATTTGCTACTCAAAAGATTAAAGTCAAGAAATTTGGATTTATGAGATA 900
842 TATGTAATTTGCTACTCAAAAGATTAAAGTCAAGAAATTTGGATTTATGAGATA 901
901 TTAAGTATGATGGAAGGCCAATAATAGCAGTTACCTTAATGTTGTCACAAATCTTA 960
902 TTAAGTATGATGGAAGGCCAATAATAGCAGTTACCTTAATGTTGTCACAAATCTTA 960
961 TTAATTAACAACACACAGCTTAATTAATGCTCAAAAGTGTACTACTTGTCAACCTCA 1020
962 TTAATTAACAACACACAGCTTAATTAATGCTCAAAAGTGTACTACTTGTCAACCTCA 1020
1021 GTTAAAGCTTTGTTGATGAGAGGTGGACCAAAATTCCTTATTAACAACAAGAT 1080
1021 GTTAAAGCTTTGTTGATGAGAGGTGGACCAAAATTCCTTATTAACAACAAGAT 1080
1081 GTGACTTTGATATTGATGACCAAAATTTGAAACAAAGCCTTACATCAGTTTAAAGG 1140
1081 GTGACTTTGATATTGATGACCAAAATTTGAAACAAAGCCTTACATCAGTTTAAAGG 1140
1141 ATCTTGATCAATGGGCAAAATTTGAAATATTTCTGAAAAGAAAGAACTGTTGAG 1200
1141 ATCTTGATCAATGGGCAAAATTTGAAATATTTCTGAAAAGAAAGAACTGTTGAG 1200
1142 ATCTTGATCAATGGGCAAAATTTGAAATATTTCTGAAAAGAAAGAACTGTTGAG 1200
1142 ATCTTGATCAATGGGCAAAATTTGAAATATTTCTGAAAAGAAAGAACTGTTGAG 1200
1201 TTTGATGTCGAAAAGTTGGCAATTAAGTCTGACCAAGCAAGAAACCGGGAGACAGCA 1260
1201 TTTGATGTCGAAAAGTTGGCAATTAAGTCTGACCAAGCAAGAAACCGGGAGACAGCA 1260
1261 TGTGAGATTCTGGAATTTTGCTTAATGTTCCAGTATGATGATTAAGTCAAGCTTCC 1320
1261 TGTGAGATTCTGGAATTTTGCTTAATGTTCCAGTATGATGATTAAGTCAAGCTTCC 1320
1321 TGTATTTGCTCTCGGGGAAACAGAGACTTCAAGGCCCCCAAGGTGACCTGGAGCTGC 1380
1321 TGTATTTGCTCTCGGGGAAACAGAGACTTCAAGGCCCCCAAGGTGACCTGGAGCTGC 1380
1381 GGGAACTCTGCTACCTGGACAACTGGTCAAGATGTAAGCTTG 1426

```

```

Db      1381 GGGAACTCTGCTACCTGGACAACTGGTCAAGATGTAAGCTTG 1426
RESULT 6
AAK53249
ID   AAK53249 standard; cDNA; 1578 BP.
XX
AC   AAK53249;
XX
CT   06-NOV-2001 (first entry);
XX
DE   Human polynucleotide SEQ ID NO 2778.
XX
KM   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM   tissue growth factor; immunomodulatory; cancer; leukaemia;
KM   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PE   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0726422.
XX
PA   (HXS-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI   Zhang CA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI   Xie M, Yang Y, Wejhrman T, Goodrich R;
XX
DR   WPI: 2001-476283/51.
XX
PT   P-PSDS; AAM80116.
XX
Nucleic acids encoding polypeptides with cytokine-like activities,
useful in diagnosis and gene therapy -
XX
Claim 1; Page 4980-4981; 6221pp; English.
XX
The invention relates to polynucleotides (AAK51456-AAK53435) and the
encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
cytokine, cell proliferation or cell differentiation or which may induce
production of other cytokines in other cell populations. The
polynucleotides and polypeptides are useful in gene therapy, vaccines or
peptide therapy. The polypeptides have various cytokine-like activities,
e.g. stem cell growth factor activity, haematopoiesis regulating
activity, tissue growth factor activity, immunomodulatory activity and
activin/inhibin activity and may be useful in the diagnosis and/or
treatment of cancer, leukaemia, nervous system disorders, arthritis and
inflammation.
CC   Note: Records for SEQ ID NO 2:10 (AAK52581), 211; (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
Sequence 1578 BP; 496 A; 294 C; 359 G; 429 T; 0 other;
Query Match      43.8%; Score 1254; DB 22; Length 1578;
Best Local Similarity 59.9%; Pred. No. 0;
Matches 1424; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Cy      1 ATGCTCAGTAAATTAATTAATTTCTGCAATGTTTGGTCTGCTTCAAGATTTCTG 60
Db      2 ATGCTCAGTAAATTAATTAATTTCTGCAATGTTTGGTCTGCTTCAAGATTTCTG 61

```



```

QY 2064 CGGATTCACGAGAAAAAGGGGACAAAGAAATCAAGTGAAGAAAGTATTCAGGCTCA 2:23
DB 489 CGGATTCAGGAGAAAAAGGGGACAAAGAAATCAAGTGAAGAAAGTATTCAGGCTCA 548
QY 2124 AAAAGGAGAAAAATGSAAGACAGGAAATTCAGAGGCAACAGGAAATTCAGAGCCATCATGG 2183
DB 549 AAAAGGAGAAAAATGSAAGACAGGAAATTCAGAGGCAACAGGAAATTCAGAGCCATCATGG 608
QY 2184 TGCAAAAGAGAGAGAGGTGAAAAAGGAGAACTGGTCCGAGAGTCCATTCGATCAAA 2243
DB 609 TGCAAAAGAGAGAGAGGTGAAAAAGGAGAACTGGTCCGAGAGTCCATTCGATCAAA 668
QY 2244 AGGAGAAATCTGGGGTGTGATGGCTTGATGAGGAGCCGACAGTCTTAAGGGGCAACCTGGGGA 2203
DB 669 AGGAGAAATCTGGGGTGTGATGGCTTGATGAGGAGCCGACAGTCTTAAGGGGCAACCTGGGGA 728
QY 2304 TCCAGGTCTCAGAGGACCCCGAGGTTTGAGTGGAGAGCCGGAAGAGATTTTCAGAGCA 2363
DB 729 TCCAGGTCTCAGAGGACCCCGAGGTTTGAGTGGAGAGCCGGAAGAGATTTTCAGAGCA 788
QY 2364 ATTATTCGACAGTTCGACAGATGTATTAAGAGCCAGCTACCACTTACTTCAGAG 2423
DB 789 ATTATTCGACAGTTCGACAGATGTATTAAGAGCCAGCTACCACTTACTTCAGAG 848
QY 2424 TGGAGAAATTTAGAAATTTGTATCATTCGCTCTCCACATGCTCCCGGGTATTCCTGG 2483
DB 849 TGGAGAAATTTAGAAATTTGTATCATTCGCTCTCCACATGCTCCCGGGTATTCCTGG 908
QY 2484 GCCACCTGCTCCGATAGGCCCGAGAGGCTCCAGAGATTAACCTGTTTGCAGAGAGAG 2543
DB 909 GCCACCTGCTCCGATAGGCCCGAGAGGCTCCAGAGATTAACCTGTTTGCAGAGAGAG 968
QY 2544 TGGTGTCTTCGATTTAGTGGGCTGCTCGAGAGCTTCAGAGTTCAGAGATTTAAAGGCTT 2603
DB 969 TGGTGTCTTCGATTTAGTGGGCTGCTCGAGAGCTTCAGAGTTCAGAGATTTAAAGGCTT 1028
QY 2604 ACCAGAGAAATTTGGGAAAAAGGAGCCAGAGGTTTGGGATCTCGAGAGCAAGGCTC 2663
DB 1029 ACCAGAGAAATTTGGGAAAAAGGAGCCAGAGGTTTGGGATCTCGAGAGCAAGGCTC 1088
QY 2664 TCCGTGTCCTCCAGGTTCAGAGGCTCTCTCGATTAAGCAAAAGAGTCTCCAGAGAG 2723
DB 1089 TCCGTGTCCTCCAGGTTCAGAGGCTCTCTCGATTAAGCAAAAGAGTCTCCAGAGAG 1148
QY 2724 CCGAGGTCTCCCTGGCAAGATGAGACCATGAGAAACCTGGAATTCAGAGGCAACGAGG 1208
DB 1149 CCGAGGTCTCCCTGGCAAGATGAGACCATGAGAAACCTGGAATTCAGAGGCAACGAGG 1268
QY 2784 CCCCCAGGATCTCGAGCCCATCTACTATGTTTATGTAATTCAGAGAGAGATCCGTT 2843
DB 1209 CCCCCAGGATCTCGAGCCCATCTACTATGTTTATGTAATTCAGAGAGAGATCCGTT 1268
QY 2844 CAGAAAGAGACCAAACTATTAG 2865
DB 1269 CAGAAAGAGACCAAACTATTAG 1290

```

RESULT 5

AAK53227 standard, cDNA, 1578 BP.

```

ID AAK53227
XX AAK53227;
XX 06-NOV-2001 (first entry)
DE Human polynucleotide SEQ ID NO 2756.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX

```

```

OS Homo sapiens.
XX
XX W0200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX
XX 27-APR-2000; 2000US-0560875.
XX
XX 20-JUN-2000; 2000US-0580075.
XX
XX 19-JUL-2000; 2000US-0620325.
XX
XX 01-SEP-2000; 2000US-0654936.
XX
XX 15-SEP-2000; 2000US-0663561.
XX
XX 20-OCT-2000; 2000US-0693325.
XX
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Dimaac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
XX Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW,
XX Xue AJ, Yang Y, Wejthman T, Goodrich R;
XX
XX W21: 2006-476283/51.
XX
XX P-PSDB; AAK8094.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX
XX Claim 1; Page 4966; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK51435) and the
XX encoded polypeptides (AAK78323-AAK80302) that exhibit activity e-acting to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAK80302) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.
XX
XX Sequence 1578 BP; 496 A; 294 C; 359 G; 429 T; 0 other;
XX
XX Query Match 43.8%; Score 1254; DB 22; Length 1578;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 1424; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
XX
QY 1 ATGGCTCACTATTATCACTTCTCGAGGTTTGGTGGCTTTCGAAATTCGTC 60
DB 2 ATGGCTCACTATTATCACTTCTCGAGGTTTGGTGGCTTTCGAAATTCGTC 61
QY 61 TTAGCTGAAGAAGGGAAGTAAAGATCAAGTTGTCTACTGCTCCGACAGATTTAGTTTC 120
DB 62 TTAGCTGAAGAAGGGAAGTAAAGATCAAGTTGTCTACTGCTCCGACAGATTTAGTTTC 121
QY 121 ATCTTAAGTGGCTCTTAAGTGTGGCCCAAAAACCTTGAATTAAGAAAAAGTGGCTT 180
DB 122 ATCTTAAGTGGCTCTTAAGTGTGGCCCAAAAACCTTGAATTAAGAAAAAGTGGCTT 181
QY 181 GTCAATATCAAAAAAACCTTGAATTAAGGAGGCGCAAGTTTATCAAGTTGAGTGGTTC 240
DB 182 GTCAATATCAAAAAAACCTTGAATTAAGGAGGCGCAAGTTTATCAAGTTGAGTGGTTC 241
QY 241 TATAGTACTACCTGTGCTGAGATTCCTTCTGGAAGCTATGATTAAGAGAAACATTTG 300
DB 242 TATAGTACTACCTGTGCTGAGATTCCTTCTGGAAGCTATGATTAAGAGAAACATTTG 301
QY 301 ACGGAGAGAGTGAATCATCTACTTGAAGAGAAACCAAAAGACAGGAGGCCATC 360

```

ABX73394
 ID ABX73394 standard; DNA; 2375 BP.
 AC ABX73394;
 DT 18-MAR-2003 (first entry)
 XX
 XX Human novel polynucleotide #222.
 XX
 XX Human; gene; ds; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 KW haemostatic; antiarteriosclerotic.
 XX
 OS Homo sapiens.
 XX
 XX US2002132753-A1.
 XX
 XX 19-SEP-2002.
 XX
 XX 17-JAN-2001; 2001US-0764864.
 XX
 XX 31-JAN-2000; 2000US-179065P.
 XX 04-FEB-2000; 2000US-180628P.
 XX 28-JUN-2000; 2000US-214886P.
 XX 07-JUL-2000; 2000US-216647P.
 XX 07-JUL-2000; 2000US-216980P.
 XX 11-JUL-2000; 2000US-217487P.
 XX 11-JUL-2000; 2000US-217496P.
 XX 14-JUL-2000; 2000US-218290P.
 XX 26-JUL-2000; 2000US-220963P.
 XX 26-JUL-2000; 2000US-220964P.
 XX 14-AUG-2000; 2000US-224518P.
 XX 14-AUG-2000; 2000US-224519P.
 XX 14-AUG-2000; 2000US-225267P.
 XX 14-AUG-2000; 2000US-225268P.
 XX 14-AUG-2000; 2000US-225270P.
 XX 14-AUG-2000; 2000US-225447P.
 XX 14-AUG-2000; 2000US-225757P.
 XX 22-AUG-2000; 2000US-225758P.
 XX 30-AUG-2000; 2000US-228688P.
 XX 01-SEP-2000; 2000US-228924P.
 XX 01-SEP-2000; 2000US-229287P.
 XX 01-SEP-2000; 2000US-229343P.
 XX 01-SEP-2000; 2000US-229345P.
 XX 05-SEP-2000; 2000US-229509P.
 XX 05-SEP-2000; 2000US-229513P.
 XX 08-SEP-2000; 2000US-231413P.
 XX 21-SEP-2000; 2000US-234223P.
 XX 21-SEP-2000; 2000US-234274P.
 XX 25-SEP-2000; 2000US-234997P.
 XX 27-SEP-2000; 2000US-235834P.
 XX 29-SEP-2000; 2000US-236327P.
 XX 29-SEP-2000; 2000US-236367P.
 XX 29-SEP-2000; 2000US-236368P.
 XX 29-SEP-2000; 2000US-236369P.
 XX 29-SEP-2000; 2000US-236370P.
 XX 02-OCT-2000; 2000US-236802P.
 XX 02-OCT-2000; 2000US-237037P.
 XX 02-OCT-2000; 2000US-237038P.
 XX 02-OCT-2000; 2000US-237039P.
 XX 02-OCT-2000; 2000US-237040P.
 XX 13-OCT-2000; 2000US-239935P.
 XX 20-OCT-2000; 2000US-240960P.
 XX 20-OCT-2000; 2000US-241785P.
 XX 20-OCT-2000; 2000US-241809P.
 XX 01-NOV-2000; 2000US-244617P.
 XX 17-NOV-2000; 2000US-249299P.
 XX 08-DEC-2000; 2000US-251856P.

08-DEC-2000; 2000US-251868P.
 08-DEC-2000; 2000US-251869P.
 (ROSE/) ROSEN C A.
 (RUBB/) RUBEN S M.
 (BARA/) BAPASH S C.
 Rosen CA, Ruben SM, Barash SC;
 WPI; 2003-147444/14.
 P-PDB; ABU55134.
 New polypeptides and nucleic acids, useful in gene therapy for
 PT treating, inhibiting or preventing e.g. neural, immune system,
 PT muscular, respiratory, reproductive, gastrointestinal, pulmonary,
 PT cardiovascular or renal disorders -
 XX
 C:aim 1: SEQ ID NO 232; 402pp; English.
 XX
 CC The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid
 CC arthritis and multiple sclerosis), muscular disorders, respiratory
 CC diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),
 CC reproductive disorders, gastrointestinal disorders, pulmonary disorders,
 CC cardiovascular disorders (e.g. congenital heart defects, Ebstein's
 CC anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute
 CC kidney failure and end-stage renal disease), hyperproliferative disorders
 CC (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g.
 CC septic shock, bursitis and appendicitis), allergic reactions and
 CC conditions (e.g. asthma), blood related disorders (e.g. thrombosis,
 CC atherosclerosis and myocardial infarction) and cancerous diseases.
 CC Sequences ABX73173-ABX74167 represent human novel polynucleotides of the
 CC invention.
 XX
 SQ Sequence 2375 BP; 731 A; 435 C; 600 G; 609 T; 0 other;
 Query Match 44.7%; Score 1282; DB 25; Length 2375;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1534 AAAGGTGAATGGTGCCCAAGGAGACAAAGGATCAGCTGGATTTTATGCGAAAGGG 1643
 DB 9 AAAGGTGAATGGTGCCCAAGGAGACAAAGGATCAGCTGGATTTTATGCGAAAGGG 69
 QY 1644 TGC AAAAGGTGAAAAGGGGAATGCTGGGTTTCCCTGCGCTCCCTGGACCTGCTGGAGAACC 1703
 DB 69 TGC AAAAGGTGAAAAGGGGAATGCTGGGTTTCCCTGCGCTCCCTGGACCTGCTGGAGAACC 128
 QY 1704 AGAAGACATGGAAGGATGGATTAAATGGGTAGTCCCGGTTTCAAGGAGAACGAGGATC 1763
 DB 129 AGAAGACATGGAAGGATGGATTAAATGGGTAGTCCCGGTTTCAAGGAGAACGAGGATC 188
 QY 1764 CCTGTGCTCCGGGCGAGATGGAACACAGGGAGAGCTGGAATCCAGGATTTCCTGG 1823
 DB 185 CCTGTGCTCCGGGCGAGATGGAACACAGGGAGAGCTGGAATCCAGGATTTCCTGG 248
 QY 1824 AAACCGAGGATTAATGGGCCAAAAGGGAGAAATTGGGCTCCAGGACAGCAAGAAAAA 1893
 DB 249 AAACCGAGGATTAATGGGCCAAAAGGGAGAAATTGGGCTCCAGGACAGCAAGAAAAA 309
 QY 1884 AGGAGCCCGAGGATGCTGTGTTTAAATGGGAGCAATGGCTACAGGCCAGCTGGAAC 1943
 DB 309 AGGAGCCCGAGGATGCTGTGTTTAAATGGGAGCAATGGCTACAGGCCAGCTGGAAC 368
 QY 1944 ACCGGATCTAAGGGAAGCAAGGTGAACCTGGGAATTCAGGGATGCTGGGGCTTCTGG 2003
 DB 369 ACCGGATCTAAGGGAAGCAAGGTGAACCTGGGAATTCAGGGATGCTGGGGCTTCTGG 428
 QY 2004 GCTCAAGGAGAACCCAGGAGCAACGGGTTCCCGAGGAGAACCCAGGATACATGGGTTTACC 2063
 DB 429 GCTCAAGGAGAACCCAGGAGCAACGGGTTCCCGAGGAGAACCCAGGATACATGGGTTTACC 488

PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-488783/53.
 DR P-PSDB; AAU16066.
 DR
 XX
 XX New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Claim 1; SEQ ID No 232; 980pp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence encodes a novel secreted protein of the invention.

Query Match 44.7%; Score 1282; DB 22; Length 2375;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1282; Conservative 0; Mismatches 0; Indels 0; Gaps 3;

QY 1584 AAAGGGTGAATGGGTGCCAAAGGAGACAAAGATCACTGGATTTTATGCAAAAGGG 1643
 DB 9 AAAGGGTGAATGGGTGCCAAAGGAGACAAAGATCACTGGATTTTATGCAAAAGGG 68
 QY 1644 TCGAAAGGTGAAGAGGGAATGCTGGCTTCCTGGCTCCCTGGACCTGGTGGAGACC 1703
 DB 69 TCGAAAGGTGAAGAGGGAATGCTGGCTTCCTGGCTCCCTGGACCTGGTGGAGACC 128
 QY 1704 AGGAAGACATGAAAGGATGGATTAATGGGTAGTCCCGGTTCCTGGAGAGAGCAGGATC 1763
 DB 129 AGGAAGACATGAAAGGATGGATTAATGGGTAGTCCCGGTTCCTGGAGAGAGCAGGATC 189
 QY 1764 CCTGTGTCTCCGGGCGAGGATGGAAACACGGGGAGAGCTCGGAATCCCGAGGATTCCTGG 1823
 DB 189 CCTGTGTCTCCGGGCGAGGATGGAAACACGGGGAGAGCTCGGAATCCCGAGGATTCCTGG 248
 QY 1824 AAACCGAGGATTAATGGGCCAAAGAGGAGAAATTTGGGCTCCAGGACACGAGGAAAAA 1893

DB 249 AAACCGAGGATTAATGGGCCAAAGAGGAGAAATTTGGGCTCCAGGACAGCAAGGAAAAA 308
 QY 1884 AGGAGCCCCCAGGAGTGCCTGGTTTAAATGGGAAGCAATGGCTCACCAGGCCAGCCTGGAC 1943
 DB 309 AGGAGCCCCCAGGAGTGCCTGGTTTAAATGGGAAGCAATGGCTCACCAGGCCAGCCTGGAC 368
 QY 1944 ACCGGGATCTAAGGGAAGCAAAAGGTGAACCTGGAATTCAGGGATGSCCTGGGGCTTCTGG 2003
 DB 369 ACCGGGATCTAAGGGAAGCAAAAGGTGAACCTGGAATTCAGGGATGSCCTGGGGCTTCTGG 428
 QY 2004 GCTCAAGGAGAACCCAGGAGCAACCGGTTCCCGAGGAGAACCCAGGATACATGGGTTTACC 2063
 DB 429 GCTCAAGGAGAACCCAGGAGCAACCGGTTCCCGAGGAGAACCCAGGATACATGGGTTTACC 488
 QY 2064 CGGGATTCAGGGAAGAAAGGGGACAAAGGAATCAAGGTGAAGAGGTATTTCAGGGTCA 2123
 DB 489 CGGGATTCAGGGAAGAAAGGGGACAAAGGAATCAAGGTGAAGAGGTATTTCAGGGTCA 548
 QY 2124 AAAGGAGAAAAATGGAAGACAGGGGAATTCAGGGCAACAGGGAAATTCAGGGCCATCATGG 2183
 DB 549 AAAGGAGAAAAATGGAAGACAGGGGAATTCAGGGCAACAGGGAAATTCAGGGCCATCATGG 608
 QY 2184 TCGAAAAGGAGAGAGGTGAAAAGGAGAACCTGGTGTCCAGAGTGCATTTGGATCAAA 2243
 DB 609 TCGAAAAGGAGAGAGGTGAAAAGGAGAACCTGGTGTCCAGAGTGCATTTGGATCAAA 668
 QY 2244 AGGAGAAATCTGGGTGGATGGTTGATGGGGCCCGCAGGTCCTAAAGGGCAACCTGGGA 2303
 DB 669 AGGAGAAATCTGGGTGGATGGTTGATGGGGCCCGCAGGTCCTAAAGGGCAACCTGGGA 728
 QY 2304 TCCAGTCTCTCAGGGACCCCGCAGGTTTGGATGGGAAGCCCGGAAGAGAGTTTTCAGAA 2363
 DB 729 TCCAGTCTCTCAGGGACCCCGCAGGTTTGGATGGGAAGCCCGGAAGAGAGTTTTCAGAA 788
 QY 2364 ATTATTCGAAAGTTTGACAGATGTATTAAGAGCCAGCTACAGTCTTACTTCAGAG 2423
 DB 789 ATTATTCGAAAGTTTGACAGATGTATTAAGAGCCAGCTACAGTCTTACTTCAGAG 848
 QY 2424 TGAAGAATTTAGAAATTTGATCATTCGCTGTCCCAACATGGTCCCGGGTATTCTCTGG 2483
 DB 849 TGAAGAATTTAGAAATTTGATCATTCGCTGTCCCAACATGGTCCCGGGTATTCTCTGG 908
 QY 2484 GCCACCTGTCTCCATAGGCCCCAGAGGTTCCAGAGGATTAACCTGGTTTCCAGGAAGAGA 2543
 DB 909 GCCACCTGTCTCCATAGGCCCCAGAGGTTCCAGAGGATTAACCTGGTTTCCAGGAAGAGA 968
 QY 2544 TGGTGTCTCTGGTTAGTGGGTGTCCTCGAGCTCCAGGTTCAGAGGATTAAGAGGCT 2603
 DB 969 TGGTGTCTCTGGTTAGTGGGTGTCCTCGAGCTCCAGGTTCAGAGGATTAAGAGGCT 1028
 QY 2604 ACCAGGAAGAAATGGGGAAAAAGGGAGCAAGGGTTTGGGTATCTCTGGAGAACAGGCTC 2663
 DB 1029 ACCAGGAAGAAATGGGGAAAAAGGGAGCAAGGGTTTGGGTATCTCTGGAGAACAGGCTC 1088
 QY 2664 TCTGTGTCTCCAGGTTCAGAGGCCCTCTCGAATTAAGCAAGAAAGGTCTCTCCAGGAGA 2723
 DB 1089 TCTGTGTCTCCAGGTTCAGAGGCCCTCTCGAATTAAGCAAGAAAGGTCTCTCCAGGAGA 1148
 QY 2724 CCCAGGTCTCTCTGGCAAGATGGAGACCATGGAAACCTGGAATCCAGGGCAACCCAGG 2783
 DB 1149 CCCAGGTCTCTCTGGCAAGATGGAGACCATGGAAACCTGGAATCCAGGGCAACCCAGG 1208
 QY 2784 CCCCCCAGGCATCTCGACCCCATCTATGTTTATGTTAATTTGCCAGAGAGATCCCGTT 2843
 DB 1209 CCCCCCAGGCATCTCGACCCCATCTATGTTTATGTTAATTTGCCAGAGAGATCCCGTT 1268
 QY 2844 CAGAAAAGGACCAACTATTAG 2865
 DB 1269 CAGAAAAGGACCAACTATTAG 1290

XX
DT
XX
DE
XX
XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;
cytostatic; cardiac; vasotropic; cerebroprotective; nootropic;
neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
vulnerary; secreted protein; rheumatoid arthritis;
hyperproliferative disorder; rheumatoid arthritis;
cerebrovascular disorder; cerebral ischaemia; angiogenesis;
nervous system disorder; Alzheimer's disease; infection; ocular disorder;
corneal infection; wound healing; epithelial cell proliferation;
skin ageing; food additive; preservative; antiproliferative.
OS
XX Homo sapiens.
XX
FN WO200155322-A2.
XX
XX
PD 02-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001WO-US01341.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232197.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 23-OCT-2000; 2000US-0241787.
PR 23-OCT-2000; 2000US-0241808.
PR 23-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 31-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.

Db 1223 GTTAAGACGTTCTTGTGTAAGGCTGGCACCAAAATTCGTCCTTAGTAAACAGAACAGAT 1282
Qy 1081 GTGACCTTTGTATATATGATGACCAACAAATTTGAAACAAAGCCCTTACATCCAGTTTATGGG 1:40
Db 1283 GTGACCTTTGTATATATGATGACCAACAAATTTGAAACAAAGCCCTTACATCCAGTTTATGGG 1342
Qy 1141 ATCTTGATCAATGGGCAACCCAAATTTGMAAATATCTGGAAGAAGAACTGTTTCAG 1200
Db 1343 ATCTTGATCAATGGGCAACCCAAATTTGMAAATATCTGGAAGAAGAACTGTTTCAG 1402
Qy 1201 TTTGATGTCCTCAAAAGTTGGAATCTTACTGTGACCCAGAACAGAACAAACCGGAGACAGCA 1260
Db 1403 TTTGATGTCCTCAAAAGTTGGAATCTTACTGTGACCCAGAACAGAACAAACCGGAGACAGCA 1462
Qy 1261 TGTGAGATCTCTGGATTT-----TGCCTTAATGCTCCAGTGATGTAGGTTCAACT 1311
Db 1463 TGTGAGATCTCTGGATTTAATGGAGAGTGCCCTTAATGCTCCAGTGATGTAGGTTCAACT 1522
Qy 1312 CCAGCTCCCTGTATTTGTCCTCCGGGAAAAACAGGACTTCAAGGCCCAAAAGTGACCCCT 1371
Db 1523 CCAGCTCCCTGTATTTGTCCTCCGGGAAAAACAGGACTTCAAGGCCCAAAAGTGACCCCT 1582
Qy 1372 GGAATGCTGGGAACCTGGCTACCTCGACAAACCTGTCAAGATGTGTAAGCTTGATAT 1431
Db 1583 GGAATGCTGGGAACCTGGCTACCTCGACAAACCTGTCAAGATGTGTAAGCTTGATAT 1642
Qy 1432 CAGGAAATTCAGGAGACACCAAGTGTTCAGGATCTCCAGGAATACAAAGAGCTCGAGGA 1491
Db 1643 CAGGAAATTCAGGAGACACCAAGTGTTCAGGATCTCCAGGAATACAAAGAGCTCGAGGA 1702
Qy 1492 CTACAGGTTTCAAAAGAGAACCCAGGGGAGATGTGACAGGGTGATCTGAGATCTCT 1551
Db 1703 CTACAGGTTTCAAAAGAGAACCCAGGGGAGATGTGACAGGGTGATCTGAGATCTCT 1762
Qy 1552 GGTTCCTCGGCTTCATGGCATGCGAGATCAAAAGGTTGAAATGGGTCCTCAAAAGGAGAC 1611
Db 1763 GGTTCCTCGGCTTCATGGCATGCGAGATCAAAAGGTTGAAATGGGTCCTCAAAAGGAGAC 1822
Qy 1612 AAAGGATCACTGGATTTTATGGCAAAAGGTTGAAAGGTTGAAAGGGAATGCTGGC 1671
Db 1823 AAAGGATCACTGGATTTTATGGCAAAAGGTTGAAAGGTTGAAAGGGAATGCTGGC 1882
Qy 1672 TTTCCCTGGCTCCCTGGACCTCTGGAGAACAGGAGACATGGAAGGAGGATTAATG 1731
Db 1883 TTTCCCTGGCTCCCTGGACCTCTGGAGAACAGGAGACATGGAAGGAGGATTAATG 1942
Qy 1732 GGTAGTCCCGTTTCAAGGAGAGCAGGATCCCTGTGTCTCCGGGACAGGATGGAACA 1791
Db 1943 GGTAGTCCCGTTTCAAGGAGAGCAGGATCCCTGTGTCTCCGGGACAGGATGGAACA 2002
Qy 1792 CGGGAGACCTTGGATCCAGGATTTCTGGAAACCCAGGATTAATGGGCCAAAGGGA 1851
Db 2003 CGGGAGACCTTGGATCCAGGATTTCTGGAAACCCAGGATTAATGGGCCAAAGGGA 2062
Qy 1852 GAAAATTGGGCTCCAGGACAGCAAGGAAAAAAGAGGCCCGGATCCCTGGTTTAATG 1911
Db 2063 GAAAATTGGGCTCCAGGACAGCAAGGAAAAAAGAGGCCCGGATCCCTGGTTTAATG 2122
Qy 1912 GGAAGCAATGGCTTACAGGACAGCAAGGAAAAAAGAGGCCCGGATTAAGGGAAGCAAGGTGAA 1971
Db 2123 GGAAGCAATGGCTTACAGGACAGCAAGGAAAAAAGAGGCCCGGATTAAGGGAAGCAAGGTGAA 2182
Qy 1972 CTGGAATTCAGGAGTGTCTGGGCTTCTGGGCTCAAGGAGAACACAGGAGACAGGGT 2031
Db 2183 CTGGAATTCAGGAGTGTCTGGGCTTCTGGGCTCAAGGAGAACACAGGAGACAGGGT 2242
Qy 2032 TCCCCAGGAGAACCCAGGATACATGGGTTTACCCGGATTTCAAGGAAAAAAGGGGACAAA 2091
Db 2243 TCCCCAGGAGAACCCAGGATACATGGGTTTACCCGGATTTCAAGGAAAAAAGGGGACAAA 2302
Qy 2092 GGAATTCAGGAGTGTCAAGGATTCAGGCTCAAAAGGAGAAATGGAAGACAGGATTT 2151
Db 2303 GGAATTCAGGAGTGTCAAGGATTCAGGCTCAAAAGGAGAAATGGAAGACAGGATTT 2362

Qy 2152 CCAGGGCAACAGGGAATTCAGGGCATCATCTGTCGAAAAAGAGAGAGAGGTGAAAAAGGA 2211
Db 2363 CCAGGGCAACAGGGAATTCAGGGCATCATCTGTCGAAAAAGAGAGAGGTGAAAAAGGA 2422
Qy 2212 GAACCTGTGTCCAGAGTGCCATTTGGATCAAAAGGAGAACTCTGGGTGGATGGCTTCATG 2271
Db 2423 GAACCTGTGTCCAGAGTGCCATTTGGATCAAAAGGAGAACTCTGGGTGGATGGCTTCATG 2482
Qy 2272 GGGCCCGCAGGTCCTTAAAGGGCAACCTGGGATCCAGTCTCTCAGGACCCCGCAGGTTG 2331
Db 2483 GGGCCCGCAGGTCCTTAAAGGGCAACCTGGGATCCAGTCTCTCAGGACCCCGCAGGTTG 2542
Qy 2332 GATGGGAAGCCCGGAAGAGATTTTCAGAACAAATTTATCGACAAATTTTCACAGATGA 2391
Db 2543 GATGGGAAGCCCGGAAGAGATTTTCAGAACAAATTTATCGACAAATTTTCACAGATGA 2602
Qy 2392 ATAGAGCCCGCAGCTACCAAGTCTTACTTCAGAGTGGAAAGAAATTTAGAAATTTGATCATTC 2451
Db 2603 ATAGAGCCCGCAGCTACCAAGTCTTACTTCAGAGTGGAAAGAAATTTAGAAATTTGATCATTC 2662
Qy 2452 CTGTCCCAACATGGCTCCCGGATTTCTTGGGCACTGTGTCGATAGGCCACAGAGGT 2511
Db 2663 CTGTCCCAACATGGCTCCCGGATTTCTTGGGCACTGTGTCGATAGGCCACAGAGGT 2722
Qy 2512 CCCAGAGGATTTACCTGTGTTTCCAGGAAGAGATGTGTCTCTGGATTTAGTGGGTGTCCCT 2571
Db 2723 CCCAGAGGATTTACCTGTGTTTCCAGGAAGAGATGTGTCTCTGGATTTAGTGGGTGTCCCT 2782
Qy 2572 GGAAGTCCAGGTCAGAGGATTTAAAGGCTTACAGGAAGAAATGGGAAAAAGGGAGC 2631
Db 2783 GGAAGTCCAGGTCAGAGGATTTAAAGGCTTACAGGAAGAAATGGGAAAAAGGGAGC 2842
Qy 2632 CAAAGGTTTGGGTATCTCTGGAGAACAGGTCCTCTGTGTCCTCCAGGTCAGAGAGGCCCT 2691
Db 2843 CAAAGGTTTGGGTATCTCTGGAGAACAGGTCCTCTGTGTCCTCCAGGTCAGAGAGGCCCT 2902
Qy 2692 CCGTGAATTAAGCAAGAGGTCCTCCAGGAGACCCAGGTCCTCTCGCAAGATGGAGAC 2751
Db 2903 CCGTGAATTAAGCAAGAGGTCCTCCAGGAGACCCAGGTCCTCTCGCAAGATGGAGAC 2962
Qy 2752 CATGAAAACCTTGAATCCAGGGCAACCCAGGCCCCCGGATCTCGGACCCCATCACTA 2811
Db 2963 CATGAAAACCTTGAATCCAGGGCAACCCAGGCCCCCGGATCTCGGACCCCATCACTA 3022
Qy 2812 TGTCTTAGTGAATTTGCCAGAGAGATCCGTTCAGAAAAGGACCAAACTATTAG 2865
Db 3023 TGTCTTAGTGAATTTGCCAGAGAGATCCGTTCAGAAAAGGACCAAACTATTAG 3076

RESULT 2

AAK52243
ID AAK52243 standard; cDNA; 1826 bp.
XX
AC AAK52243;
XX
DT 06-NOV-2001 (first entry)
XX
DE Humar. polynucleotide SEQ ID NO 788.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
CS Homo sapiens.
XX
PN W0200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX

C 82 19 0.7 2787 23 AAS77413 DNA encoding novel
 C 83 19 0.7 2787 23 AAS77593 DNA encoding novel
 C 84 19 0.7 2787 23 AAS82312 DNA encoding novel
 C 85 19 0.7 2845 18 AAT93093 Human GCFl cDNA.
 C 86 19 0.7 3188 23 AB119112 Drosophila melanog
 C 87 19 0.7 3423 20 AAX40183 Lung cancer associ
 C 88 19 0.7 3523 18 AAT93091 Human transcriptio
 C 89 19 0.7 3600 22 AAK85958 Human immune/haema
 C 90 19 0.7 3608 25 AB118028 Aspergillus fumiga
 C 91 19 0.7 3609 25 AB119842 Aspergillus fumiga
 C 92 19 0.7 6741 22 AAK85961 Human immune/haema
 C 93 19 0.7 8766 24 ABN72581 Streptococcus agal
 C 94 19 0.7 10437 22 AAS41709 Genomic sequence #
 C 95 19 0.7 12620 22 ABA16691 Human nervous syst
 C 96 19 0.7 48000 22 AAF27996 Human calcium sens
 C 97 19 0.7 66685 22 AAS07390 Human genomic DNA
 C 98 19 0.7 66686 24 ABS73149 Human CLASP-5 geno
 C 99 19 0.7 160755 23 AAH88704 Human DNA sequenc
 C 100 19 0.7 910715 20 AAX20248 Borrelia burgdorfe

ALIGNMENTS

RESULT 1
 ABX71234
 ID ABX71234 standard; cDNA; 4160 BP.

AC ABX71234;
 DT 15-APR-2003 (first entry)

Cell structure and mobility-associated cDNA from clone DKFZphfbr2_2b5.

Human; gene; gene therapy; vaccine; disease treatment; detection; ss.

Homo sapiens.

W0200112659-A2.

22-FEB-2001.

18-AUG-2000; 2000MO-IB01496.

18-AUG-1999; 99US-0149499.

28-SEP-1999; 99US-0156503.

(GEHU-) GERMAN HUMAN GENOME PROJECT.

Wiemann S;

WPI; 2001-327840/34.

P-PSDB; ABUS2683.

Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies -

Claim 1; Page 201-202; 1095pp; English.

This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence encodes a polypeptide described in the disclosure of the invention.

Sequence 4160 BP; 1271 A; 795 C; 1026 G; 1068 T; 0 other;

Query Match 71.9%; Score 2061; DB 23; Length 4160;

Best Local Similarity 99.5%; Pred. No. 0;
 Matches 2861; Conservative 0; Mismatches 4; Indels 9; Gaps 1;
 QY 2 ATGGCTCCTATATACATTTCTCTGCTGTTTGGTGGTCTCTTCTCAGAAATTTCTGTG 60
 DB 203 ATGGCTCCTATATACATTTCTCTGCTGTTTGGTGGTCTCTTCTCAGAAATTTCTGTG 262
 QY 61 TTAGCTGAAGATGGGAAGTAAGATCAAGTCTCTGCTGCTGCTCCGACAGATTTAGTTTTC 120
 DB 263 TTAGCTGAAGATGGGAAGTAAGATCAAGTCTCTGCTGCTCCGACAGATTTAGTTTTC 322
 QY 121 ATCTTAGATGGCTTTATAGTGTGGCCGAGAAACTTTGAAATAGTCAAAAAGTGGTTC 180
 DB 323 ATCTTAGATGGCTTTATAGTGTGGCCGAGAAACTTTGAAATAGTCAAAAAGTGGTTC 382
 QY 181 GTCATATATCAGAAAACCTTTGACATAGCGCGAGTATTATCAAGTTGGAGTGGTTC 240
 DB 383 GTCATATATCAGAAAACCTTTGACATAGCGCGAGTATTATCAAGTTGGAGTGGTTC 442
 QY 241 TATAGTGACTTACCTGCTGCTGGAGATTCCTCTCGGAAGCTATGATTCAGGAGAAATTTG 300
 DB 443 TATAGTGACTTACCTGCTGCTGGAGATTCCTCTCGGAAGCTATGATTCAGGAGAAATTTG 502
 QY 301 ACGGACAGCTGGATCCATCTACTTAGGAGAAACACAAAGACAGGAGGAGCCATC 360
 DB 503 ACGGACAGCTGGATCCATCTACTTAGGAGAAACACAAAGACAGGAGGAGCCATC 562
 QY 361 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCTCAGCACTTTCTGACTAAGATAGCAGTG 420
 DB 563 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCTCAGCACTTTCTGACTAAGATAGCAGTG 622
 QY 421 GTACTTACGATGGGAAATCCCAAGATGACGTCAAGGATGAGCTCAAGCAGCAAGAGAT 480
 DB 623 GTACTTACGATGGGAAATCCCAAGATGACGTCAAGGATGAGCTCAAGCAGCAAGAGAT 682
 QY 481 AGTAAGATAACATTTTGTCTATTTGGTGTGGTTCAGAAAACAGAAAGATGCGCAACTAGA 540
 DB 683 AGTAAGATAACATTTTGTCTATTTGGTGTGGTTCAGAAAACAGAAAGATGCGCAACTAGA 742
 QY 541 GCTATTGCCAAACAGCTTTCGCTACTTATGTTTATGTTGGAAGACTATATTGCAATA 600
 DB 743 GCTATTGCCAAACAGCTTTCGCTACTTATGTTTATGTTGGAAGACTATATTGCAATA 802
 QY 601 TCCAAATATAGGAGTGTGATGAAGCAGAACTTTCTGAGAGATCTCTGTCCACACAGA 660
 DB 803 TCCAAATATAGGAGTGTGATGAAGCAGAACTTTCTGAGAGATCTCTGTCCACACAGA 862
 QY 661 ATTCCAGTGGCAGCTCGTGATGAAAAGGGGATTTGATATTTCTTTAGGTTAGATGTAAT 720
 DB 963 ATTCCAGTGGCAGCTCGTGATGAAAAGGGGATTTGATATTTCTTTGGGTTAGATGTAAT 922
 QY 721 AAAAAGGTTAAGAAAGATACAGCTTTCACCAAAAAGATTAAGAGATATGAAGTAACA 780
 DB 923 AAAAAGGTTAAGAAAGATACAGCTTTCACCAAAAAGATTAAGAGATATGAAGTAACA 982
 QY 781 TCAAAAGTTGATTTATCAGAACTCAAGCAATGTTTCCCAAGAGGTTCTTCCCTCAATCA 840
 DB 983 TCAAAAGTTGATTTATCAGAACTCAAGCAATGTTTCCCAAGAGGTTCTTCCCTCAATCA 1042
 QY 841 TATGTTATTTGCTCTACTCAAGATTTAAAGTCAAGAAAATTTGGGATTTATGGAGAATA 900
 DB 1043 TATGTTATTTGCTCTACTCAAGATTTAAAGTCAAGAAAATTTGGGATTTATGGAGAATA 1102
 QY 901 TTAACATTTATGGAAGGCCCAAAATACAGTTACCTTTAAATGTTGGAGCAAAATCTTA 960
 DB 1103 TTAACATTTATGGAAGGCCCAAAATACAGTTACCTTTAAATGTTGGAGCAAAATCTTA 1162
 QY 961 TTAATTTAACCAACAGCGGTAAATTAATGGCTCACAGTGGTTTACCTTTGGCTTAACCTCAA 1020
 DB 1163 TTAATTTAACCAACAGCGGTAAATTAATGGCTCACAGTGGTTTACCTTTGGCTTAACCTCAA 1222
 QY 1021 GTTAAGACGTTTGTGATGAGGCTGGCAACCAATTCCTCTCTTAGTAACAGAAAGAT 1080

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 19:26:03 ; Search time 715 Seconds
(without alignments)
10816.629 Million cell updates/sec

Title: US-09-996-611b-5
Perfect score: 2865
Sequence: 1 atgggtcattattacatt.....gaaaggaccacatttag 2865

Scoring table: OLIGO NUC
Gapop_60.0, Capext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database: N Geneseq 19Jun03:*

1:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
25:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2061	71.9	4160	23	ABX71234
2	1426	49.8	1826	22	AAK52243
3	1282	44.7	2375	22	AAK526053
4	1282	44.7	2375	25	ABX73394
5	1254	43.8	1578	22	AAK53227
6	1254	43.8	1578	22	AAK53249
7	1203	42.0	1691	22	AAK52265
8	1062	37.1	2230	22	AAK52045

Result No.	Score	Query Match	Length	ID	Description
1	2061	71.9	4160	23	ABX71234
2	1426	49.8	1826	22	AAK52243
3	1282	44.7	2375	22	AAK526053
4	1282	44.7	2375	25	ABX73394
5	1254	43.8	1578	22	AAK53227
6	1254	43.8	1578	22	AAK53249
7	1203	42.0	1691	22	AAK52265
8	1062	37.1	2230	22	AAK52045

9	1062	37.1	2230	22	AAI58402
10	974	34.0	1184	24	ABQ72640
11	966	33.7	2209	22	AAK53029
12	966	33.7	2209	22	AAI60188
13	923	32.2	1169	24	ABQ72511
14	749	26.1	1115	24	AAK31132
15	382	13.3	1414	24	ABX71703
16	252	8.8	447	20	AAV86274
17	166	5.8	221	22	AAI49752
18	131	4.6	246	21	AAK29851
19	82	2.9	181	22	ABK69546
20	63	2.2	1101	24	ABK71706
21	28	1.0	466	22	AAK36639
22	22	0.8	2154	22	AAK81046
23	20	0.7	393	21	AAK43283
24	20	0.7	393	24	ABZ14620
25	20	0.7	479	22	ABA57633
26	20	0.7	479	22	AAK05687
27	20	0.7	479	22	AAK31302
28	20	0.7	479	22	AAK37199
29	20	0.7	479	23	ABS30983
30	20	0.7	479	24	ABS06054
31	20	0.7	548	20	AAZ08316
32	20	0.7	548	21	AAK79168
33	20	0.7	548	23	AAZ23243
34	20	0.7	577	24	ABU83527
35	20	0.7	638	21	AAK38427
36	20	0.7	954	22	AAF92514
37	20	0.7	1287	23	AAK51375
38	20	0.7	1299	23	AAK53120
39	20	0.7	1635	20	AAK13558
40	20	0.7	1635	24	ABS99353
41	20	0.7	1817	23	AAK90001
42	20	0.7	2493	24	ABK34875
43	20	0.7	2706	23	ABU16061
44	20	0.7	4757	23	ABU16060
45	20	0.7	4453	20	AAK23519
46	20	0.7	267156	24	ABU68560
47	19	0.7	330	22	AAK33364
48	19	0.7	380	22	AAK33364
49	19	0.7	380	22	AAK93365
50	19	0.7	387	22	AAK62758
51	19	0.7	484	21	AAK34237
52	19	0.7	573	23	ABU18649
53	19	0.7	834	22	ABZ20496
54	19	0.7	902	22	AAK06575
55	19	0.7	1023	11	AAQ06119
56	19	0.7	1072	16	AAQ84357
57	19	0.7	1074	16	AAQ84356
58	19	0.7	1137	16	AAQ84355
59	19	0.7	1140	11	AAQ05372
60	19	0.7	1143	24	ABK66364
61	19	0.7	1146	24	ABU46717
62	19	0.7	1174	22	AAK323110
63	19	0.7	1300	22	AAK46059
64	19	0.7	1300	25	AAK57817
65	19	0.7	1300	25	ABK98287
66	19	0.7	1300	25	ABK98789
67	19	0.7	1300	25	AAK05834
68	19	0.7	1300	25	ABK97878
69	19	0.7	1300	25	ABK78662
70	19	0.7	1300	25	ABK75675
71	19	0.7	1300	25	ABK76880
72	19	0.7	1300	25	ABK16720
73	19	0.7	1312	10	AAK90033
74	19	0.7	1363	24	ABZ16067
75	19	0.7	1477	16	AAQ84354
76	19	0.7	1744	22	ABA20240
77	19	0.7	2182	22	AAH16271
78	19	0.7	2341	24	AAK94897
79	19	0.7	2458	23	ABU19290
80	19	0.7	2573	23	ABU18648
81	19	0.7	2575	11	AAQ04026

Encodes ...

JOURNAL Genome Res. 10 (7), 939-949 (2000)
 MEDLINE 20359837
 PUBMED 10899143
 REFERENCE 3 (bases 1 to 681)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/tetraodon.
 FEATURES
 source
 1. .681
 /organism="Tetraodon nigroviridis"
 /mol_type="genomic DNA"
 /db_xref="taxon:9983"
 /clone="144M06"
 /clone_lib="G"
 /note="Genoscope sequence ID : COAG144BG03LPl-end : T7"
 BASE COUNT 128 a 181 c 188 g 183 t 1 others
 ORIGIN
 Query Match 0.8%; Score 22; DB 29; Length 681;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1763 CCCCTGGTGCTCCGGGGCAGGA 1784
 |||||
 Db 308 CCCCTGGTGCTCCGGGGCAGGA 329

Search completed: October 3, 2003, 04:32:35
 Job time : 5327 secs

```

RESULT 43
AV592493/c
LOCUS
DEFINITION AV592493 Bos taurus cartilage fetus Bos taurus cDNA clone
E1CA001H05 3', mRNA sequence.
ACCESSION AV592493
VERSION AV592493.1 GI:9707650
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 471)
Takauega,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE 21570554
PUBMED 1171328
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@ccoc.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.
FEATURES
source
1..471
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="E1CA001H05"
/tissue_type="cartilage"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus cartilage fetus"
/notes="vector: pZLI; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT 78 a 155 c 121 g 113 t 4 others
ORIGIN
Query Match 0.8% Score 22; DB 9; Length 471;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CQ 2659 GGTCTCTCTGTCCTCCCGAGTC 2680
|||||
Db 169 GGTCTCTCTGTCCTCCCGAGTC 148

RESULT 44
BH882288
LOCUS
DEFINITION BH882288.1 WGS-Zmay5F (JM107 adapted methyl filtered) Zea mays
genomic clone hw36d12 5', genomic survey sequence.
ACCESSION BH882288
VERSION BH882288.1 GI:22118185
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 485)
Rabinowicz,P.D., O'Shaughnessy,A.L., Balijs,V., Dedhia,N.,
Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,

```

```

Zutavern,T., McCombie,W.R. and Martienssen,R.A.
Genomic shotgun sequences from Zea mays (methyl-filtered)
Unpublished
Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cehl.org
Plate: hw36 row: d column: 12
Seq primer: -21M13UnivFwd
Class: shotgun
High quality sequence stop: 485.
FEATURES
Location/Qualifiers
1..485
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="hw36d12"
/lab_host="JM107 or DH5a"
/clone_lib="WGS-Zmay5F (JM107 adapted methyl filtered)"
/notes="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
The vector was digested with XbaI and one nucleotide was
added by fill in in the recessive 3' end. The genomic DNA
was religated, end repaired, adaptor ligated and size
fractionated using sephadex. The resulting fragments were
between 0.8 and 3 kb and were cloned into the vector
i: x/y reads in M13mp19, b/g reads in pUC19. The same
ligation was transformed in either JM107 or DH5a."
BASE COUNT 188 a 77 c 86 g 134 t
ORIGIN
Query Match 0.8% Score 22; DB 28; Length 485;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CQ 753 AAAAAGATATAAGGATATGAA 774
|||||
Db 226 AAAAAGATATAAGGATATGAA 247

RESULT 45
CNS02JXM
LOCUS
DEFINITION CNS02JXM 681 bp DNA linear GSS 01-SEP-2000
Tetraodon nigroviridis genome survey sequence 17 end of clone
194M06 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL208803
VERSION AL208803.1 GI:7859148
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
REFERENCE 1
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
JOURNAL 20296633
MEDLINE 10835645
PUBMED
AUTHORS
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Fizames,C., Fischer,C., Bouneau,L., Billaud,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

```

MEDLINE 21570554
PUBMED 11713328
COMMENT

Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@cocoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

FEATURES

source
1..440
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="E1AD003D11"
/cell_type="an adipocyte cell line"
/lab_host="DH10B"
/clone_lib="Bos taurus adipocyte cell line"
/note="Vector: pZ11; Site 1: SalI; Site 2: NotI; Poly A
was deleted from a NotI site"

BASE COUNT 80 a 148 c 102 g 110 t

ORIGIN

Query Match 0.8%; Score 22; DB 9; Length 440;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GGTCTCTCTGTCCTCCAGGTC 2680

Db 383 GGTCTCTCTGTCCTCCAGGTC 362

RESULT 41

BH882289/c

LOCUS BH882289 450 bp DNA linear GSS 05-AUG-2002
DEFINITION hw36dl2.g1 WGS-ZmaysF (JM107 adapted methyl filtered) Zea mays
genomic clone hw36dl2 5', genomic survey sequence.

ACCESSION BH882289

VERSION BH882289.1

KEYWORDS GSS

SOURCE Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 450)

AUTHORS Rabinowicz, P.D., O'Shaughnessy, A.L., Ballja, V., Dedhia, N.,
Katzemburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,
Zutavern, I., McCombie, W.R., and Martienssen, R.A.
Genomic shotgun sequences from Zea mays (methyl-filtered);
Unpublished

TITLE

JOURNAL

COMMENT

Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874

Email: mcombie@cshl.org

Plate: hw36 row: d column: 12

Seq primer: -21M31UnivFwd

Class: shotgun

High quality sequence stop: 487.

FEATURES

source
1..450
/organism="Zea mays"
/mol_type="genomic DNA"
/culturivar="B73"
/db_xref="taxon:4577"
/clone="hw36dl2"
/lab_host="JM107 or DH5a"
/clone_lib="WGS-ZmaysF (JM107 adapted methyl filtered)"

/note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;
The vector was digested with XbaI and one nucleotide was
added by fill in in the recessive 3' end. The genomic DNA
was nebulized, end repaired, adaptor ligated and size
fractionated using sephadex. The resulting fragments were
between 0.8 and 3 kb and were cloned into the vector
(.x/y reads in M13:p19, .b/g reads in pUC19). The same
ligation was transformed in either JM107 or DH5a."

BASE COUNT 119 a 79 c 70 g 182 t

ORIGIN

Query Match 0.8%; Score 22; DB 28; Length 450;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 753 AAAAAGATATAAAGGATATGAA 774

Db 245 AAAAAGATATAAAGGATATGAA 224

RESULT 42

AV599587/c

LOCUS AV599587 454 bp mRNA linear EST 27-NOV-2001
DEFINITION AV599587 Bos taurus cartilage fetus Bos taurus cDNA clone
E1CA044H10 3', mRNA sequence.

ACCESSION AV599587

VERSION AV599587.1

KEYWORDS EST

SOURCE Bos taurus (cow)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 454)

AUTHORS Takasuga, A., Hirotsune, S., Itoh, R., Jitchazon, A., Suzuki, H., Aso, H.
and Sugimoto, Y.

TITLE Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs

JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)

MEDLINE 21570554

PUBMED 11713328

COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division

Shirakawa Institute of Animal Genetics

Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan

Tel: 81-248-25-5641

Fax: 81-248-25-5725

Email: kazusugi@cocoa.ocn.ne.jp

Single pass sequencing.

This clone was obtained from a polyA-deleted cDNA library.

FEATURES

source

1..454
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="E1CA044H10"
/tissue_type="cartilage"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus cartilage fetus"
/note="Vector: pZ11; Site 1: SalI; Site 2: NotI; Poly A
was deleted from a NotI site"

BASE COUNT 78 a 154 c 115 g 107 t

ORIGIN

Query Match 0.8%; Score 22; DB 9; Length 454;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GGTCTCTCTGTCCTCCAGGTC 2680

Db 192 GGTCTCTCTGTCCTCCAGGTC 161

Email: kazuugi@cocoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

FEATURES
source Location/Qualifiers

1..367
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="E1CA023A01"
/tissue_type="cartilage"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus cartilage fetus"
/note="Vector: pZHI; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"

BASE COUNT 65 a 124 c 88 g 89 t
ORIGIN

Query Match 0.8%; Score 22; DB 9; Length 367;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GGTCCTCTGGTCCCGGAGTC 2680

DB 191 GGTCCTCTGGTCCCGGAGTC 160

RESULT 38

LOCUS CB801379

DEFINITION CB801379 412 bp mRNA linear EST 16-MAY-2003
AMGNNUC:MRPE4-00162-D9-A mrpe4 (10380) Rattus norvegicus cDNA clone
mrpe4-00162-d9 5', mRNA sequence.

ACCESSION CB801379

VERSION 1 GI:29914582

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 412)

AUTHORS Amgen EST Program.

TITLE Amgen Rat EST Program

JOURNAL Unpublished

COMMENT Contact: Dan Fitzpatrick

Amgen, Inc

One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA

Tel: 805 447-4881

Plate: 00162 row: d column: 9.

FEATURES
source Location/Qualifiers

1..412
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="mrpe4-00162-d9"
/tissue_type="placenta embryo"
/clone_lib="mrpe4 (10380)"
/note="Vector: pSPOR1; Site_1: SalI; Site_2: NotI;
placenta embryo day 17"

BASE COUNT 83 a 122 c 147 g 60 t

ORIGIN

Query Match 0.8%; Score 22; DB 14; Length 412;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2702 GCAAGAAGGTCCTCCAGGAGA 2723

DB 315 GCAAGAAGGTCCTCCAGGAGA 336

RESULT 39

AV597967/c

LOCUS AV597967

DEFINITION AV597967 424 bp mRNA linear EST 27-NOV-2001
E1CA035B11 3', mRNA sequence.

ACCESSION AV597967

VERSION AV597967.1 GI:9715461

KEYWORDS EST.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 424)

AUTHORS Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.

TITLE Establishment of a high throughput EST sequencing system using

poly(A) tail-removed cDNA libraries and determination of 36,000

bovine ESTs

Nucleic Acids Res. 29 (22), E108 (2001);

MEDLINE 21570554

PubMed 11713328

COMMENT

Contact: Yoshikazu Sugimoto

Shirakawa Institute of Animal Genetics

Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan

Tel: 81-248-25-5641

Fax: 81-248-25-5725

Email: kazuugi@cocoa.ocn.ne.jp

Single pass sequencing.

This clone was obtained from a polyA-deleted cDNA library.

FEATURES
source Location/Qualifiers

1..424

/organism="Bos taurus"

/mol_type="mRNA"

/db_xref="taxon:9913"

/clone="E1CA035B11"

/tissue_type="cartilage"

/dev_stage="fetus"

/lab_host="DH10B"

/clone_lib="Bos taurus cartilage fetus"

/note="Vector: pZHI; Site_1: SalI; Site_2: NotI; Poly A

was deleted from a NotI site"

BASE COUNT 74 a 144 c 109 g 96 t 1 others

ORIGIN

Query Match 0.8%; Score 22; DB 9; Length 424;

Best Local Similarity 100.0%; Pred. No. 27;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GGTCCTCTGGTCCCGGAGTC 2680

DB 185 GGTCCTCTGGTCCCGGAGTC 164

RESULT 40

LOCUS AV614172/c

DEFINITION AV614172 440 bp mRNA linear EST 28-NOV-2001
E1A0003D11 3', mRNA sequence.

ACCESSION AV614172

VERSION AV614172

KEYWORDS EST.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 440)

AUTHORS Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.

TITLE Establishment of a high throughput EST sequencing system using

poly(A) tail-removed cDNA libraries and determination of 36,000

bovine ESTs

Nucleic Acids Res. 29 (22), E108 (2001);

JOURNAL

Qy 1405 CCTGGTCAAGATGTAAGCTCGGATATCAGGAAT 1439
 |||||||
 Db 346 CCTGGTCAAGATGTAAGCTCGGATATCAGGAAT 380
 |||||||

RESULT 35
 AQ514534 556 bp DNA linear GSS 05-MAY-1999
 LOCUS
 DEFINITION HS 5081_B1_A01_SP6E RPCI-11 Human Male BAC Library Homo sapiens
 genomic_clone Plate=657 Col=1 Row=B, genomic survey sequence.

ACCESSION AQ514534
 VERSION AQ514534.1 GI:4742898
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 556)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 PUBMED 10449764
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieteradejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
 or from Resear h Genetics (info@resgen.com). BAC end Web Server:
 http://www.htsc.washington.edu
 Plate: 657 row: B column: 1
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 556.
 Location/Qualifiers
 1..556
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="Plates:657 Col=1 Row=B"
 /sex="male"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /note="vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI Methylase. Size selected DNA was cloned into the
 pBACe3.6 vector at EcoRI sites" 5 others

BASE COUNT 191 a 106 c 103 g 151 t
 ORIGIN

Query Match 0.8%; Score 23; DB 28; Length 556;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1833 ATTAATGGCCAAAGGAGAAA 1855
 |||||||
 Db 367 ATTAATGGCCAAAGGAGAAA 389
 |||||||

RESULT 36
 CC086742
 LOCUS
 DEFINITION CSU-K33r.15N22.SP6 CSU-K33r Aedes aegypti genomic clone

ACCESSION CC086742
 VERSION CC086742.1 GI:29940194
 KEYWORDS GSS.
 SOURCE Aedes aegypti (yellow fever mosquito)
 ORGANISM Aedes aegypti

REFERENCE 1 (bases 1 to 980)
 AUTHORS Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
 TITLE End sequencing of Aedes aegypti BACs
 JOURNAL Unpublished
 COMMENT Other_GSSs: CSU-K33r.15N22.T7
 Contact: Brendan Loftus
 Department of Eukaryotic Genomics
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-3543
 Fax: 301-838-0208
 Email: entas@tigr.org
 Library was provided by Susan Brown and Dennis Knudson at Colorado
 State University.
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1..980
 /organism="Aedes aegypti"
 /mol_type="genomic DNA"
 /strain="Rexville"
 /db_xref="taxon:7159"
 /clone="CSU-K33r.15N22"
 /clone_lib="CSU-K33r"
 /note="Vector: pBelOBAC11; Site 1: HindIII"
 BASE COUNT 305 a 198 c 204 g 273 t
 ORIGIN

Query Match 0.8%; Score 23; DB 29; Length 980;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 429 GGATGGCAATCCCAAGATGACG 451
 |||||||
 Db 367 GGATGGCAATCCCAAGATGACG 945
 |||||||

RESULT 37
 AV595980/c
 LOCUS
 DEFINITION AV595980 Bos taurus cartilage fetus Bos taurus cDNA clone
 E1CA023A01.3, mRNA sequence.
 ACCESSION AV595980
 VERSION AV595980.1 GI:9712989
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus

REFERENCE 1 (bases 1 to 367)
 AUTHORS Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
 and Sugimoto,Y.
 TITLE Establishment of a high throughput EST sequencing system using
 poly(A) tail-removed cDNA libraries and determination of 36,000
 bovine ESTs
 JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)
 MEDLINE 21570554
 PUBMED 11713328
 COMMENT Contact: Yoshikazu Sugimoto
 Animal Genetics Division
 Shirakawa Institute of Animal Genetics
 Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
 Tel: 81-248-25-5641
 Fax: 81-248-25-5725

```

/db_xref="taxon:9606"
/clone="Plate=1003 Col=13 Row=F"
/sex="male"
/clone_lib="RRCI-11 Human Male BAC Library"
/note="vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT      217 a  124 c   96 g   213 t   18 others
ORIGIN
Query Match      1.3%; Score 36; DB 28; Length 668;
Best Local Similarity 100.0%; Pred.No. 2.2e-06;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1135 TTAGGATCTTGATCAATGGCGCAACCCAAATTTGGA 1173
|||||
Db 323 TTAGGATCTTGATCAATGGCGCAACCCAAATTTGGA 288
|||||

RESULT 33
BM724875
LOCUS      691 bp      mRNA      linear      EST 01-MAR-2002
DEFINITION UI-E-E01-aja-a-05-0-UI-r1 UI-E-E01 Homo sapiens cDNA clone
ACCESSION  BM724875
VERSION     BM724875
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 691)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
PubMed      8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 Reverse.
            Location/Qualifiers
                1..691
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="UI-E-E01-aja-a-05-0-UI"
                /tissue_type="fetal eye"
                /dev_stage="fetal"
                /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                /clone_lib="UI-E-E01"
                /note="Organ: eye; Vector: p773-Pac (Pharmacia) with a
                modified polylinker; Site 1: EcoRI; Site 2: Not I;
                UI-E-E01 is a normalized cDNA library containing the
                following tissue(s): fetal eye. The library was
                constructed according to Bonaldo, Lennon and Soares,
                Genome Research, 6:791-806, 1996. First strand cDNA
                synthesis was primed with an oligo-dT primer containing a
                Not I site. Double stranded cDNA was ligated to an EcoRI I
                adaptor, digested with Not I, and cloned directionally

```

```

into p773-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
GCGGTATACC. This library was created for the program, Gene
Discovery in the Visual System, supported by National Eye
Institute (NEI)."
BASE COUNT      206 a  120 c  131 g   233 t    1 others
ORIGIN
Query Match      1.3%; Score 36; DB 12; Length 691;
Best Local Similarity 100.0%; Pred.No. 2.3e-06;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2830 AGAAGAGATCCGTTCCAGAAAGGACCAACTATTAG 2865
|||||
Db 1 AGAAGAGATCCGTTCCAGAAAGGACCAACTATTAG 36
|||||

RESULT 34
BE899929
LOCUS      531 bp      mRNA      linear      EST 25-APR-2001
DEFINITION 180456 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION  BE899929
VERSION     BE899929.1 GI:10387679
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
ORGANISM    Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 531)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.C., Stone,R.T.,
Cassas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-McKown,C.G.,
Perteza,G., Holt,I., Karamycheva,S., Jiang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL     Genome Res. 11 (4), 626-630 (2001)
MEDLINE     21180013
PubMed      11282978
COMMENT     Contact: Smith TP;
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAACAGCTATGACCAT
            BACKWARD: GTTTCACGTCACGACG
            Plate: 85 row: E column: 7
            Seq primer: ATTAGGTGACACTATAG.
            Location/Qualifiers
                1..531
                /organism="Bos taurus"
                /mol_type="mRNA"
                /db_xref="taxon:9913"
                /tissue_type="pooled"
                /lab_host="DH10B"
                /clone_lib="MARC 4BOV"
                /note="Vector: pCMV SPOR6; Site 1: NotI; Site 2: SalI;
                Library made from pooled tissue from day 20 and day 40
                embryos."
BASE COUNT      161 a  120 c  146 g  104 t
ORIGIN
Query Match      1.2%; Score 35; DB 10; Length 531;
Best Local Similarity 100.0%; Pred.No. 6.9e-06;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

AQ356449/c
 LOCUS BX391778 Homo sapiens genomic clone 2541D7, linear GSS 24-JAN-1999
 DEFINITION CITBI-E1-2541D7.TF CITBI-E1 Homo sapiens genomic clone 2541D7, genomic survey sequence.
 ACCESSION AQ356449
 VERSION AQ356449.1 GI:4183622
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 778)
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and Venter,J.C.
 TITLE Use of BAC End Sequences from CalTech Libraries for Sequence-Ready Map Building
 JOURNAL Unpublished
 COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbe@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.htm.
 Seq primer: M13-21
 Class: BAC ends.

FEATURES

Location/Qualifiers

1..778
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="2541D7"
 /sex="male"
 /cell_type="sperm"
 /clone_lib="CITBI-E1"
 /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI; CalTech Human BAC Library D"
 289 a 109 c 118 g 262 t

BASE COUNT

ORIGIN

Query Match 1.5%; Score 42; DB 28; Length 778;
 Best Local Similarity 100.0%; Pred.No. 2e-09;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCTCACTATATACATTCTCTGCATGGTTTGGTGCTG 42
 |||||

Db 61 ATGGCTCACTATATACATTCTCTGCATGGTTTGGTGCTG 20
 |||||

RESULT 31
 BX391778/c
 LOCUS BX391778 Homo sapiens mRNA linear EST 13-MAY-2003
 DEFINITION BX391778 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0D1071YH18 3-PRIME, mRNA sequence.
 ACCESSION BX391778
 VERSION BX391778.1 GI:30607719
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 883)
 AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqreg@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 3377.r For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BA1027ZH02_CS02512_1&cluster=3377.r.
 Contact : Feng Liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/InvitrogenCorporation.1600>
 Faraday Avenue Genoscope sequence ID : CS0BA1027ZH02_CS02512_1.

FEATURES

Location/Qualifiers

1..883
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1071YH18"
 /tissue="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(GT) primer. Five prime end enriched double-strand cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."
 274 a 166 c 275 t 5 others

BASE COUNT

ORIGIN

Query Match 1.3%; Score 38; DB 13; Length 883;
 Best Local Similarity 100.0%; Pred.No. 2.3e-07;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2828 CCAGAGAGATCCGTTCCAGAAAGGACCAAACTATTAG 2865
 |||||

Db 777 CCAGAGAGATCCGTTCCAGAAAGGACCAAACTATTAG 740
 |||||

RESULT 32
 AQ836646/c
 LOCUS AQ836646 Homo sapiens DNA linear GSS 30-AUG-1999
 DEFINITION HS 5427 B1 C07 57A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1003 Col=13 Row=F, genomic survey sequence.
 ACCESSION AQ836646
 VERSION AQ836646.1 GI:5906520
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 668)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc.Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 PUBMED 13449764
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98105, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (<http://www.htsc.washington.edu>).
<http://www.htsc.washington.edu>
 Plate: 1003 row: F column: 13
 Seq primer: 7
 Class: BAC ends
 High quality sequence stop: 668.

FEATURES

Location/Qualifiers

1..668
 /organism="Homo sapiens"
 /mol_type="genomic DNA"

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKF2p686B2113"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlc3)"
/notes="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"
BASE COUNT 193 a 77 c 107 g 223 t
ORIGIN
Query Match 1.8%; Score 51; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 4.9e-14;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1804 GGAATCCCGAGGATTTCTGGAAACCGAGGATTAATGGGCCAAAGGGAGAA 1854
|||||
DB 532 GGAATCCCGAGGATTTCTGGAAACCGAGGATTAATGGGCCAAAGGGAGAA 582
|||||

RESULT 28
BE236040 518 bp mRNA linear EST 10-JUL-2000
LOCUS
DEFINITION
ACCESSION BE236040
VERSION
KEYWORDS
SOURCE Sus scrofa (pig)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 518)
AUTHORS
Fahrenkrug, S.C., Smith, T.P.L., Freking, B.A., Cho, J., White, J.,
Vallet, J., Wise, T., Rohrer, G.A., Perle, G., Sultana, R., Quackenbush,
J. and Keele, J.W.
Pig gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
Mamm. Genome 13 (8), 475-478 (2002)
2213789
12226715
CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68903-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCGAGTCAGGACG
Plate: 85 row: G column: 4
Seq primer: ATTAGTGCACACTAG.
FEATURES
Location/Qualifiers
1..518
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 1P1G"
/notes="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 157 a 105 c 175 g 81 t
ORIGIN
Query Match 1.6%; Score 45; DB 10; Length 518;
Best Local Similarity 100.0%; Pred. No. 5.4e-11;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1692 TGCTGGAGACCGAGACATGGGAAGGATGGATTATGGGTAG 1736
|||||
DB 111 TGCTGGAGAACCGAGAGACATGGGAAGGATGGATTATGGGTAG 155
|||||

RESULT 29
R77744 372 bp mRNA linear EST 07-JUN-1995
LOCUS
DEFINITION
ACCESSION Y180902.s1 Soares placenta Nb2HP Homo sapiens cDNA clone
VERSION Y180902.s1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 372)
AUTHORS
Hillier, L., Clark, N., Dubucq, T., Elliston, K., Hawkins, M., Holman,
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tac, F., Trevaskis, E., Waterston,
R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 850, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 1108
High quality sequence stops: 328
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyI not found
Insert Length: 1108 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 328.
FEATURES
Location/Qualifiers
1..372
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:566422"
/db_xref="taxon:9606"
/clone="IMAGE:145586"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares placenta Nb2HP"
/notes="Organ: placenta; Vector: p773D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
AACTGGAGAAATCGCGCCGAGGAATTTTITTTTITTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Patima Ronaldo. "
BASE COUNT 113 a 88 c 52 g 115 t 4 others
ORIGIN
Query Match 1.5%; Score 44; DB 14; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1576 CCAGGATCAAGGGTGAATGGTGCCAAAGGAGACAAGGATC 1619
|||||
DB 173 CCAGGATCAAGGGTGAATGGTGCCAAAGGAGACAAGGATC 130
|||||

RESULT 30

```


adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."

BASE COUNT 53 a 59 c 77 g 52 t
 ORIGIN
 Query Match 2.0%; Score 58; DB 9; Length 241;
 Best Local Similarity 100.0%; Pred.No.1e-17;
 Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2443 GATCATTTGCTTCCCAACATGGCTCCCGGGTATTCCTGGCCACCTGTCGCGATAG 2500
 |||||
 Db 1 GATCATTTGCTTCCCAACATGGCTCCCGGGTATTCCTGGCCACCTGTCGCGATAG 58

RESULT 25
 CB268816 428 bp mRNA linear EST 20-FEB-2003
 LOCUS 1007722 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
 DEFINITION cDNA 5', mRNA sequence.

ACCESSION CB268816.1 GI:28443401
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Yang, R.-Z., Shuldiner, A. and Gong, D.-W.
 TITLE EST analysis of human adipose gene expression
 JOURNAL Unpublished

COMMENT Contact: Gong Da-Wei
 Division of Endocrinology, Diabetes and Nutrition
 University of Maryland
 660 Redwood St, HH497, Baltimore, MD 21201, USA
 Tel: 410 706 1672
 Fax: 410 706 1622

Email: dgong@medicine.umaryland.edu
 PCR Primers
 FORWARD: CTCGGGAAGCGGCCATTGTGTGTT
 BACKWARD: AATACGACTACTATAGGCGCAATTGG
 Seq primer: GTTGTACCCGGGAATTC.

FEATURES
 source
 1..428
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /sex="Male and Female"
 /tissue type="Adipose"
 /clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"
 /note="Vector: lambdaTriplex"

BASE COUNT 122 a 83 c 68 g 135 t 20 others
 ORIGIN
 Query Match 2.0%; Score 57; DB 14; Length 428;
 Best Local Similarity 100.0%; Pred.No.3.9e-17;
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2809 CTATGTTTGTAGTAAATTCAGAGAGATCCGTTCCAGAAAGGACCAACTATTAG 2865
 |||||
 Db 12 CTATGTTTGTAGTAAATTCAGAGAGATCCGTTCCAGAAAGGACCAACTATTAG 68

RESULT 26
 BX387812 903 bp mRNA linear EST 08-MAY-2003
 LOCUS 1007722 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
 DEFINITION cDNA 5', mRNA sequence.

ACCESSION BX387812.1 GI:30461772
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens

ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 903)
 Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 3377.r For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0A1071DD09QPL&cluster=3377.r. Contact :

Peng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0A1071DD09QPL.
 Location/Qualifiers
 1..903
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1071H18"
 /tissue type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES
 source

294 a 139 c 173 g 258 t 39 others
 BASE COUNT
 ORIGIN
 Query Match 2.0%; Score 56; DB 13; Length 903;
 Best Local Similarity 100.0%; Pred.No.1.5e-16;
 Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2705 AAGAAGTCTCCAGGAGACCCAGGCTCTCCCTGCAGAGATGGAGACCTGGAAAA 2765
 |||||
 Db 57: AAGAAGTCTCCAGGAGACCCAGGCTCTCCCTGCAGAGATGGAGACCTGGAAAA 626

RESULT 27
 AL602341
 LOCUS

DEFINITION DKF2p686B2113_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
 DKF2p686B2113_5', mRNA sequence.

ACCESSION AL602341
 VERSION AL602341.1 GI:15165847
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Ansong, W., Winkner, U., Mewes, W., Weil, B. and Wiemann, S.
 TITLE EST (Ansong, W., Winkner, U., Mewes, H.W., Weil, B. and Wiemann, S.)
 JOURNAL Unpublished
 COMMENT Contact: Ansong W
 MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by EMBL (European Molecular Biology Laboratories,
 Heidelberg/Germany) within the cDNA sequencing consortium of the
 German Genome Project.
 No 5' sequence available.

This clone (DKF2p686B2113) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
 source

1..600

Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=st2-CM1-NN1006-300
 500-231-901&t3=2000-05-30&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 7
 High quality sequence stop: 329.

FEATURES

Location/Qualifiers
 1..330
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /db_xref="taxon:9606"
 /clone_lib="NN1006"

/note="Organ: Lung; Vector: p7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR 1; Site 2: Not 1; UI-CF-FNO is a subtracted cDNA library derived from two normal Human lung epithelial cell libraries (EN1 and D01). The library was subtracted according to according to Bionaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. For additional information, contact: bent-soares@uiowa.edu
 TAG_LIB=UI-CF-FNO
 TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
 TAG_SEQ=GGCTGTAGGC"

BASE COUNT 94 a 74 c 64 g 98 t

ORIGIN

Query Match 2.3%; Score 66; DB 10; Length 330;
 Best Local Similarity 100.0%; Pred. No. 9.4e-22;
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2747 GAGACATGGAAACCTGGATCCAAAGGCGCAACAGGCCCGCCCGAGGATCTGCGACCCAT 2806
 |||
 DB 15 GAGACATGGAAACCTGGATCCAAAGGCGCAACAGGCCCGCCCGAGGATCTGCGACCCAT 74

QY 2807 CACTAT 2812

DB 75 CACTAT 80

RESULT 23

CB241974/c

LOCUS

DEFINITION CB241974 246 bp mRNA linear EST 12-FEB-2003
 UI-CF-FNO-aga-o-08-0-UI-s1 UI-CF-FNO Homo sapiens cDNA clone

UI-CF-FNO-aga-o-08-0-UI 3', mRNA sequence.

ACCESSION CB241974

VERSION CB241974.1

KEYWORDS GI:28363618

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL Normalization and subtraction: two approaches to facilitate gene

MEDLINE

PUBMED

COMMENT

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: McCray, PB

University of Iowa

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com) or from Open Biosystems

(www.openbiosystems.com).

Seq primer: M13 FORWARD

POLYA=Yes.

FEATURES

Source

1..246

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone_lib="UI-CF-FNO-aga-o-08-0-UI"

/tissue type="Human Lung Epithelial cells"

/lab host="CHICB (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-CF-FNO"

/note="Organ: Lung; Vector: p7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR 1; Site 2: Not 1; UI-CF-FNO is a subtracted cDNA library derived from two normal Human lung epithelial cell libraries (EN1 and D01). The library was subtracted according to according to Bionaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. For additional information, contact: bent-soares@uiowa.edu
 TAG_LIB=UI-CF-FNO
 TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
 TAG_SEQ=GGCTGTAGGC"

BASE COUNT 56 a 53 c 39 g 98 t

ORIGIN

Query Match 2.2%; Score 64; DB 14; Length 246;

Best Local Similarity 100.0%; Pred. No. 9.1e-21;

Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1590 GATCAAGGCTGAATGGTCCCAAGGAGACAAAGGATCACCTGGATTTATGGCAAAA 1639
 |||
 DB 219 GATCAAGGCTGAATGGTCCCAAGGAGACAAAGGATCACCTGGATTTATGGCAAAA 160

QY 1640 AGGG 1643

DB 159 AGGG 156

RESULT 24

AA247907

LOCUS

DEFINITION

AA247907 241 bp mRNA linear EST 11-MAR-1997

AA247907 56, mRNA sequence.

AA247907.1 GI:1880395

EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CDNAs from human fetal heart (1997)

Unpublished

Contact: Liew CC

Brigham and Women's Hospital

Harvard Medical School

75 Francis St. Boston, MA 02115, USA

Tel: 6177328915

Fax: 6179750995

Email: cliw@rics.bwh.harvard.edu

PCR Primers

FORWARD: 5' OCCAAGCTCGAAATTAACCTCTCACTAAAGG 3'

BACKWARD: 5' CCAGTGAATTGTAATAGCACTACTATAGGCG 3'

Seq primer: 5' GAAATTAACCTCTCACTAAAGG 3'.

FEATURES

Source

1..241

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/lab host="E. coli XL1-Blue"

/clone_lib="Human fetal heart, Lambda Zap Express"

/note="Vector: Lambda Zap Express; Site 1: EcoRI; Site 2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dT

AUTHORS
Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A.,
Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
TITLE
Expressed sequence tag analysis of human RPE/choroid for the
NEBank Project: Over 6000 non-redundant transcripts, novel genes
and splice variants
JOURNAL
Mol. Vis. 8 (4), 205-220 (2002)
MEDLINE
22103460
PUBMED
12107410
COMMENT
Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 89 row: c column: 07
Seq primer: M13RP1 reverse primer (AB1).
FEATURES
source
1. .531
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="cs99c07"
/tissue_type="RPE/choroid"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Human Retinal pigment epithelium/choroid cDNA
(Un-normalized, unamplified): cs"
/note="Organ: Eye; Vector: pCMVSPORT6; Two different doncr
eyes (75-80 years old) yielded approximately 600 mg of
dissected RPE/choroid tissue. This in turn yielded 340 ug
of total RNA and 7 ug of mRNA. A directionally cloned cDNA
library in the pCMVSPORT6 vector was constructed at Life
Technologies (Rockville, MD; now part of Invitrogen Corp),
essentially following the protocols of the SuperScript
Plasmid System (Invitrogen Corp).
<http://www.invitrogen.com/> The library code
designation was cs. For this library, cDNA inserts were
cloned into the NotI/MluI sites of the vector. EST
analysis was performed on the unamplified library at the
NIH Intramural Sequencing Center (NISC)."
BASE COUNT 149 a 87 c 107 g 188 t
ORIGIN
Query Match 3.0%; Score 85; DB 14; Length 531;
Best Local Similarity 100.0%; Pred. No. 2.2e-31;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2195 AGAGAGGTGAAAGGGAGAACCTGGTGTCCGAGGTGCCATTGGATCAAAAGGAGATCTG 2254
|||||
DB 447 AGAGAGGTGAAAGGGAGAACCTGGTGTCCGAGGTGCCATTGGATCAAAAGGAGATCTG 506
QY 2255 GGGTGGATGGCTTGTGATGGGGCCCGC 2279
|||||
DB 507 GGGTGGATGGCTTGTGATGGGGCCCGC 531
RESULT 21
LOCUS R58412 161 bp mRNA linear EST 02-MAY-1996
DEFINITION G3577 Fetal heart Homo sapiens cDNA clone G3577 5' end, mRNA
sequence.
ACCESSION R58412
VERSION R58412.1 GI:828470
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 161)
AUTHORS Hwang,D.M., Fung,Y.W., Wang,R.X., Laurensen,C.M., Ng,S.H., Lam
W.Y., Tsui,K.W., Fung,K.P., Waye,M., Lee,C.Y. and Liew,C.C.
Analysis of expressed sequence tags from a fetal human heart cDNA

JOURNAL
Genomics 30 (2), 293-298 (1995)
MEDLINE
96163883
PUBMED
8586430
COMMENT
Contact: Liew CC
Brigham and Women's Hospital
Harvard Medical School
75 Francis St. Boston, MA 02115, USA
Tel: 6177328915
Fax: 6179750995
Email: cliew@rics.bwh.harvard.edu
Seq primer: GTGGCGACGACTCTCGAGCC.
Location/Qualifiers
1. .161
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="G3577"
/lab_host="E. coli Y1090"
/clone_lib="Fetal heart"
/note="Vector: Lambda gt22; Site 1: NotI; Site 2: SalI;
mRNA was purified from human fetal hearts (10-12 weeks).
cDNA was constructed using a NotI-Oligo dT adaptor-primer.
SalI adaptors were ligated, followed by digestion with
NotI, for direction cloning into predigested lamda gt22.
Method is described in J. Mol. Cell. Cardiol. (1994) 26,
1329-1333"
BASE COUNT 51 a 29 c 50 g 31 t
ORIGIN
Query Match 2.6%; Score 74; DB 14; Length 161;
Best Local Similarity 100.0%; Pred. No. 6.5e-26;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1854 GGAATCCAGGATTTCTGGAAACCGAGGATTATGCGCCAAAGGAGAAATGGGCTT 1863
|||||
DB 52 GGAATCCAGGATTTCTGGAAACCGAGGATTATGCGCCAAAGGAGAAATGGGCTT 111
QY 1864 CCAGGACAGCAAGG 1877
|||||
DB 112 CCAGGACAGCAAGG 125
RESULT 22 6
BE701852 330 bp mRNA linear EST 12-SEP-2000
LOCUS CM1-NN1006-300500c-231-gc1 NN1006 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE701852
ACCESSION BE701852
VERSION BE701852.1 GI:13089594
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 330)
AUTHORS Dias Neto,E., Garcia Correa,R., Veklovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Garvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunsteit,A., deoliveira,P.S., Eucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE
20202663
PUBMED
10737800
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil.
Tel: +55-11-2704922
Fax: +55-11-2707001

```

Best Local Similarity 100.0%; Pred. No. 4.1e-42;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2132 AAAATGAAGACAGGAAATCCAGGCAACAGGAAATTCAGGCAATCATGTCGCAAAAG 2:91
|||||
Db 1 AAAATGAAGACAGGAAATCCAGGCAACAGGCAACAGGAAATTCAGGCAATCATGTCGCAAAAG 60
|||||

QY 2192 GAGAGAGGTTGAAAGGAGAACTCGTGTCCGAGGTGCCATTGG 2217
|||||
Db 61 GAGAGAGGTTGAAAGGAGAACTCGTGTCCGAGGTGCCATTGG 106
|||||

RESULT 19
AA451616 485 bp mRNA linear EST 05-JUN-1997
LOCUS Zx43b04.t1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone
DEFINITION IMAGE:789199 5' similar to TR:G292352 G292352 COLLAGEN CHAIN RH ;,
mRNA sequence.
ACCESSION AA451616
VERSION AA451616.1 GI:2165285
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 485)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,
T., Waterston, R., and Wilson, R.
TITLE WashU-Merck EST Project 1997
JOURNAL Unpublished
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 422.
FEATURES
Location/Qualifiers
source
1..485
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:5986816"
/db_xref="taxon:9606"
/clone="IMAGE:789199"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/clone_lib="Soares total_fetus Nb2HF8_9w"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer '5'
TGTTACCAATCTGAGTGGAGCGCGCTTAATTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaudo."
BASE COUNT 160 a 86 c 91 g 148 t
ORIGIN
Query Match 3.6%; Score 104; DB 9; Length 485;
Best Local Similarity 100.0%; Pred. No. 4.4e-41;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2397 AGCCAGCTACAGTCTTACTTACAGTGGAGAAATTGAAATTTGATCATTCCTGTC 2456
|||||
Db 64 AGCCAGCTACAGTCTTACTTACTTACAGTGGAGAAATTGAAATTTGATCATTCCTGTC 123
|||||

```

```

QY 2457 CCAACATGGCTCCCGGGTATTCCTGGGCCACCTGTCCTCGATAG 2503
|||||
Db 124 CCAACATGGCTCCCGGGTATTCCTGGGCCACCTGTCCTCGATAG 167
|||||

RESULT 19
N85186 387 bp mRNA linear EST 01-APR-1996
LOCUS J2812F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
DEFINITION clone J2812 5' similar to REPETITIVE ELEMENT, mRNA sequence.
ACCESSION N85186
VERSION N85186.1 GI:1260811
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 387)
AUTHORS Liew, C.C.
TITLE cDNAs from fetal heart (1996)
JOURNAL Unpublished
COMMENT Contact: Liew CC
Brigham and Women's Hospital
Harvard Medical School
75 Francis St. Boston, MA 02115, USA
Tel: 6177328915
Fax: 6179750995
Email: cliew@rics.bwh.harvard.edu
Seq primer: GAAATTAACCTCTCACTAAAGG.
FEATURES
Location/Qualifiers
source
1..387
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="J2812"
/lab_host="E. coli XLI-Blue"
/clone_lib="Human fetal heart, Lambda ZAP Express"
/notes="Vector: Lambda ZAP Express; Site1: EcoRI; Site 2:
XhoI; mRNA was purified from human fetal hearts (8-10
weeks). cDNA was synthesized using a XhoI-Oligo dT
adaptor-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI, for directional cloning into
predigested lambda ZAP Express."
BASE COUNT 123 a 73 c 126 g 65 t
ORIGIN
Query Match 3.0%; Score 85; DB 14; Length 387;
Best Local Similarity 100.0%; Pred. No. 2e-31;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2:33 AAATGGAAGACAGGGAATTCAGGCAACAGGGAATTCAGGCCATCATGTCGCAAAAGG 2192
|||||
Db 122 AAATGGAAGACAGGGAATTCAGGCAACAGGGAATTCAGGCCATCATGTCGCAAAAGG 181
|||||

QY 2193 AGAGAGAGGTGAAAAGGAGAACCT 2217
|||||
Db 182 AGAGAGAGGTGAAAAGGAGAACCT 206
|||||

RESULT 20
CA397287 531 bp mRNA linear EST 06-NOV-2002
LOCUS CS89c07.y1 Human Retinal pigment epithelium/cioroid cDNA cs89c07
DEFINITION (un-normalized, unamplified): cs Homo sapiens cDNA clone cs89c07
5', mRNA sequence.
ACCESSION CA397287
VERSION CA397287.1 GI:24734456
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 531)

```

Best Local Similarity 100.0%; Pred. No. 7.2e-54;
 Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2153 CAGGCAACAGGGAATTCAAGGCATCATGTGCAAAAGGAGAGAGAGTGAAGAGGAG 2212
 |||||
 Db 22 CAGGCAACAGGGAATTCAAGGCATCATGTGCAAAAGGAGAGAGAGTGAAGAGGAG 81
 |||||

QY 2213 AACCTGGTGTCCGAGTCCCATTCGATCAAAAGGAGATCTGGGTCGATGGCTTGATGG 2272
 |||||
 Db 82 AACCTGGTGTCCGAGTCCCATTCGATCAAAAGGAGATCTGGGTCGATGGCTTGATGG 141
 |||||

QY 2273 GCGCCGAC 2281
 |||||
 Db 142 GCGCCGAC 150
 |||||

RESULT 16
 CB852087/c
 LOCUS
 DEFINITION
 CB852087 792 bp mRNA linear EST 22-APR-2003
 UI-CF-FNO-aep-m-17-0-UI.s1 UI-CF-FNO Homo sapiens cDNA clone
 UI-CF-FNO-aep-m-17-0-UI 3', mRNA sequence.

ACCESSION
 CB852087
 CB852087.1 GI:30046979
 EST.
 SOURCE
 Homo sapiens (human)

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS
 TITLE
 Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL
 MEDLINE
 PUBMED
 COMMENT

Genome Res. 6 (9), 791-806 (1996)
 97044477
 8889548
 Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul-mccray@iowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com) or from Open Biosystems
 (www.openbiosystems.com).
 Seq primer: M13 FORWARD

POLYA=No. Location/Qualifiers
 1..792
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-FNO-aep-m-17-0-UI"
 /tissue_type="Human Lung Epithelial cells"
 /lab_host="DH10B (life technologies) (T1 phage resistant)"
 /clone_lib="UI-CF-FNO"
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-CF-FNO is a subtracted cDNA library derived from two
 normalized Human lung epithelial cell libraries (EN1 and
 DU1). The library was subtracted according to according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. For additional information, contact:
 bento-soares@uiowa.edu
 TAG SEQ=None found"

BASE COUNT 245 a 174 c 121 g 252 t
 ORIGIN

Query Match 3.9%; Score 112; DB 14; Length 792;

Best Local Similarity 100.0%; Pred. No. 4.2e-45;
 Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1532 AGGGTGATCGTGGACTCTCTGGTCTTCTGGCTTCATGCGATGCCAGGATCAAGGGTG 1591
 |||||
 Db 390 AGGGTGATCGTGGACTCTCTGGTCTTCTGGCTTCATGCGATGCCAGGATCAAGGGTG 241
 |||||

QY 1592 AAATGGTGCCCAAGGAGACAAAGGATCACCCTGGATTTTATGGCAAAAGGG 1643
 |||||
 Db 240 AAATGGTGCCCAAGGAGACAAAGGATCACCCTGGATTTTATGGCAAAAGGG 189
 |||||

RESULT 17
 AA329319
 LOCUS
 DEFINITION
 AA329319 452 bp mRNA linear EST 20-APR-1997
 EST32913 Embryo, 12 week I Homo sapiens cDNA 5' end similar to
 similar to collagen, alpha family, mRNA sequence.

ACCESSION
 AA329319
 AA329319.1 GI:1981563
 EST.
 SOURCE
 Homo sapiens (human)

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS
 TITLE
 Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence

JOURNAL
 MEDLINE
 PUBMED
 COMMENT

Nature 377 (6547 suppl), 3-174 (1995)
 96026280
 7566098
 Other ESTs: THC180973
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hgi.html>)
 Seq primer: M13 Reverse.

Location/Qualifiers
 1..452
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="ATCC (inhost):130967"
 /db_xref="taxon:9606"
 /dev_stage="embryo, 12 wks"
 /clone_lib="Embryo, 12 week I"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI"

FEATURES
 source

BASE COUNT 126 a 90 c 132 g 101 t 3 others
 ORIGIN

Query Match 3.7%; Score 106; DB 9; Length 452;

/clone lib="Soares fetal liver spleen lNFLS"
 /note="Organ: Liver and Spleen; Vector: pT7T3 (Pharmac.a)
 with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dn) primer
 [5' - ACTGAGAGTAATTAAGATCTTTTCTTTTCTTTTCTTTT - 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmac.a), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fátima Bonaldo."
 93 a 132 c 96 q i32 t 11 others
 BASE COUNT

ORIGIN	Query Match	5.4%	Score 156	DB 14	Length 464
BR22	Best Local Similarity	100.0%	Pred. No. 1.3e-67		
	Matches 156	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1928	CAGCGGACCTGGAAACACCGGGATCTAAAGGGAAGCAAGAGGTGAACCTGGGATTCAAAGGGA	1987		
DB	204	CAGCGGACCTGGAAACACCGGGATCTAAAGGGAAGCAAGAGGTGAACCTGGGATTCAAAGGGA	145		
QY	1988	TGCCTGGGGGCTTCTGGGCTCAAGGGAGAACACAGSAGCAACGGSTTCCCCAGGAGAACCG	2047		
DB	144	TGCCTGGGGGCTTCTGGGCTCAAGGGAGAACACAGSAGGACACGGGTTCCTCCAGGAGAACCG	85		
QY	2048	GATCATGGGTTTACCCGGGATTCAGGAAAAAAGG	2083		
DB	84	GATCATGGGTTTACCCGGGATTCAGGAAAAAAGG	49		

04 081ACAT1069111ACCCCGGATTCACGGGGGAAAAAG 9 43

RESULT 15
N86630
LOCUS
DEFINITION
J9414F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
clone J94.4 5' similar to COLLAGEN TYPE XII; ALPHA-1, mRNA
sequence.
357 bp mRNA linear EST 01-APR-1996

ACCESSION
N86630
VERSION
N86630.1 GI:2439832
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 357)
Liew, C.C.
cDNAs from fetal heart (1996)
AUTHORS
Unpublished
JOURNAL
Contract: View CC
COMMENT

Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates;
Catarrhini; Hominidae; Homo.

Harvard Medical School
75 Francis St. Boston, MA 02115, USA
Tel: 617/728915
Fax: 617/9750995
Email: cliew@rics.bwh.harvard.edu
Seq primer: GAAATTAAACCTCACTAAAGG.

```

/Clone.i.b="Human fetal heart, Lambda Zap Express"
/notes="Vector: Lambda Zap Express; Site.1: EcoRI; Site.2:
XhoI; mRNA was purified from human fetal hearts (8-10
weeks). cDNA was synthesized using a XhoI-Oligo dT
adaptor-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI for directional cloning into
Predigested lambda Zap Express."
BASE COUNT: 96 a 69 c 116 g 74 t
ORIGIN
Query Match 4.5%; Score 129; DB 14; Length 357;

```

I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J., and Quackenbush, J.
 Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
 Unpublished
 Contact: John Quackenbush
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 3528
 Fax: 301 838 0208
 Email: johnq@igr.org
 Plate: 176
 Seq primer: Reverse.

FEATURES
 source
 1. 494
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="MAGE resequences, MAGE"
 /note="Vector: pBluescriptSKm"
 BASE COUNT 156 a 100 c 137 g 101 t
 ORIGIN

Query Match 5.7%; Score 164; DB 10; Length 494;
 Best Local Similarity 99.2%; Pred. No. 1.1e-71;
 Matches 264; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2133 AAATGAACACAGGGAAATCCAGGGCAACAGGGAATTCAGGCCCATCATGTCGCAAAAGG 2152
 DB 1 AAATGAACACAGGGAAATCCAGGGCAACAGGGAATTCAGGCCCATCATGTCGCAAAAGG 60
 QY 2193 AGAGAGAGTGAAGAGGAGAACCTGGTTCGAGGTGCCATGGATCAAAAGGAGAAATC 2252
 DB 61 AGAGAGAGTGAAGAGGAGAACCTGGTTCGAGGTGCCATGGATCAAAAGGAGAAATC 120
 QY 2253 TGGGGTGGATGGCTTCATCGCGCCGACCTGCTAAGGGGCAACCTGGGGATCCAGGTCC 2312
 DB 121 TGGGGTGGATGGCTTCATCGCGCCGACCTGCTAAGGGGCAACCTGGGGATCCAGGTCC 180
 QY 2313 TCAGGACCCCCAGGTTTGGATGGGAAGCCCGAAGAGAGATTTTCAGAACAAATTTATTCG 2372
 DB 181 TCAGGACCCCCAGGTTTGGATGGGAAGCCCGAAGAGAGATTTTCAGAACAAATTTATTCG 240
 QY 2373 ACAGTTTCACAGATCTAATAAGAG 2398
 DB 241 ACAAGTTTCACAGATCTAATAAGAG 266

RESULT 12
 BG699698
 LOCUS
 DEFINITION 744 bp mRNA linear EST 07-MAY-2001
 602681559F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:481419 5';
 mRNA sequence.
 ACCESSION BG699698
 VERSION BG699698.1 GI:13968272
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 744)
 NIH-MGC http://img.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

http://image.lnln.gov
 Plate: L1AM10709 row: e column: 16
 High quality sequence stop: 596.
 Location/Qualifiers
 1. 744

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:4814199"
 /tissue_type="hippocampus"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_95"
 /note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3', size-selected for average insert size 2.5 kb and normalized to ROF 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 202 a 206 c 177 g 159 t
 ORIGIN

Query Match 5.7%; Score 163; DB 10; Length 744;
 Best Local Similarity 100.0%; Pred. No. 4.1e-71;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGGCTCTACTATATTACATTTCTCTGGATGGTTTGGTCTCTTCAGAAATTCGTG 60
 DB 193 ATGGCTCTACTATATTACATTTCTCTGGATGGTTTGGTCTCTTCAGAAATTCGTG 252
 QY 61 TTAGCTGAAGATGGGAAAGTAAGATCAAGTTGCTGCTACTGCTCCGACAGATTTAGTTTC 120
 DB 253 TTAGCTGAAGATGGGAAAGTAAGATCAAGTTGCTGCTACTGCTCCGACAGATTTAGTTTC 312
 QY 221 ATCTTAGATGCTCTTATAGTTTGGCCCAAGAACTTTGAAA 163
 DB 313 ATCTTAGATGCTCTTATAGTTTGGCCCAAGAACTTTGAAA 355

RESULT 13
 AA095402
 LOCUS
 DEFINITION 161 bp mRNA linear EST 25-OCT-1996
 14401.Seq.F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
 5', mRNA sequence.
 ACCESSION AA095402
 VERSION AA095402.1 GI:1641019
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 161)
 cDNAs from fetal heart (1996)
 Liaw, C.C.
 Unpublished
 Contact: Liaw CC
 Brigham and Women's Hospital
 Harvard Medical School
 75 Francis St. Boston, MA 02115, USA
 Tel: 6177328915
 Fax: 6179750995
 Email: cliaw@rics.bwh.harvard.edu
 PCR Primers
 FORWARD: 5' CCCAAGCTCGAANTTAACCTCACTAAAGG 3'
 BACKWARD: 5' CCAAGTAATTGTAATGACTCACTATAGGGG 3'
 Seq primer: 5' GAAATTAACCTCACTAAAGG 3'.
 Location/Qualifiers
 1. 161
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 327)
AUTHORS Haller, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: T7
High quality sequence stop: 274.

FEATURES
source
1..327
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3896080"
/db_xref="taxon:9606"
/clone="IMAGE:277704"
/sex="male"
/tissue_type="multiple sclerosis lesions"
/dev_stage="Age 46"
/lab_hosts="DH10B (ampicillin resistant)"
/note="Vector: p77T3D (Pharmacia) with a modified
polylinker V type: phagemid. Site 1: Not I. Site 2: Eco R;
: 1st strand cDNA was primed with a Not I - oligo(3T)
primer [5',
TGTTACCAATCTGAGTGGGAGCGCGCATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco R;
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p77T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH)."
BASE COUNT 82 a 80 c 100 g 64 t 1 others
ORIGIN
Query Match 8.6%; Score 245; DB 14; Length 327;
Best Local Similarity 100.0%; Pred. No. 5.1e-113; Mismatches 0; Indels 0; Gaps 0;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2501 CCCAGAGGGTCCCAGAGGATTACCTGGTTTCCAGGAAGAGATGGTGTCTCTGGATTAG 2560
DB 74 GCCCAGAGGGTCCCAGAGGATTACCTGGTTTCCAGGAAGAGATGGTGTCTCTGGATTAG 133
QY 2561 TGGGTGTCCTGGAGCTCCAGCTGCAGAGATTAAAGGCTTACCAGGAGAAATGGG 2620
DB 134 TGGGTGTCCTGGAGCTCCAGCTGCAGAGATTAAAGGCTTACCAGGAGAAATGGG 193
QY 2621 AAAAAGGGAGCAAGGGTTTGGGTATCTCTGGAGAACAAAGTCTCTCTGGTCCCCCAGGTC 2680
DB 194 AAAAAGGGAGCAAGGGTTTGGGTATCTCTGGAGAACAAAGTCTCTCTGGTCCCCCAGGTC 253
QY 2681 CAGAGGGCTCTCTGGATATAGCAAGAGTCTCTCCAGGAGACCCAGGTCCTCTCTGGCA 2740
DB 254 CAGAGGGCTCTCTGGATATAGCAAGAGTCTCTCCAGGAGACCCAGGTCCTCTCTGGCA 313
QY 2741 AAGAT 2745
DB 314 AAGAT 318

RESULT 8
AA227783
LOCUS
DEFINITION
AA227783.1 Soares NHMPU S1 Homo sapiens cDNA clone IMAGE:667543 5'
similar to SW:CAIF_HUMAN Q07092 COLLAGEN ALPHA 1(XVI) CHAIN
PRECUSOR. 1; mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1..233
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:5561510"
/db_xref="taxon:9606"
/clone="IMAGE:667543"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_hosts="DH10B"
/clote_lib="Soares NHMPU S1"
/note="Organ: mixed (see below); Vector: p77T3D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 26232-26523,
340488-345479, and 484488-489479."
BASE COUNT 65 a 63 c 67 g 38 t
ORIGIN
Query Match 8.1%; Score 233; DB 9; Length 233;
Best Local Similarity 100.0%; Pred. No. 6.1e-107; Mismatches 0; Indels 0; Gaps 0;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2599 GGCCTACCAAGGAAGTGGGAAAAGGGAGCCAGGGTTGGGTATCTCTGGAGACAA 2658
DB 1 GGCCTACCAAGGAAGTGGGAAAAGGGAGCCAGGGTTGGGTATCTCTGGAGACAA 60
QY 2659 GGTCTCTCTGTCTCCCGAGTCCAGAGGCCCTCTCTGGAATAGCAAGAGTCTCTCCA 2718
DB 61 GGTCTCTCTGTCTCCCGAGTCCAGAGGCCCTCTCTGGAATAGCAAGAGTCTCTCCA 120
QY 2719 GGAGACCCAGGTCTCTCCCTGGCAAGATGGAGACCATGGAACCTGGAATCCAGGGCAA 2778
DB 121 GGAGACCCAGGTCTCTCCCTGGCAAGATGGAGACCATGGAACCTGGAATCCAGGGCAA 160
QY 2779 CCAGGCCCCCGAGGCATCTCGACCCCATCACTATGTTTATGTGTAATTGCCAG 2831

Tel.: 301 402 3452
 Fax: 301 496 0078
 Email: graeme@helix.nih.gov
 Plate: 68 row: a column: 05
 Seq primer: M13RP1 reverse primer (ABI).
 Location/Qualifiers
 1..527

FEATURES

source

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="cs68a05"
 /tissue_type="RPE/choroid"
 /dev_stage="Adult"
 /lab_host="EMDH108"
 /clone_lib="Human Retinal pigment epithelium/choroid cDNA
 (Un-normalized, unamplified): cs"
 /note="Organ: Eye; Vector: pCMVSPORT6; Two different donor
 eyes (75-80 years old) yielded approximately 600 mg of
 dissected RPE/choroid tissue. This in turn yielded 340 ug
 of total RNA and 7 ug of mRNA. A directionally cloned cDNA
 library in the pCMVSPORT6 vector was constructed at Life
 Technologies (Rockville, MD; now part of Invitrogen Corp),
 essentially following the protocols of the SuperScript
 Plasmid System (Invitrogen Corp).
 <http://www.invitrogen.com/>". The library code
 designation was cs. For this library, cDNA inserts were
 cloned into the NotI/MluI sites of the vector. EST
 analysis was performed on the unamplified library at the
 NIH Intramural Sequencing Center (NISC)."
 145 a 118 c 126 g 138 t

BASE COUNT 145 a 118 c 126 g 138 t
 ORIGIN
 Query Match 13.2%; Score 377; DB 14; Length 527;
 Best Local Similarity 100.0%; Pred. No. 2.9e-180;
 Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTCACTATATACATTTCTCTGCATGGTTTGGTGGCTGCTTTCAGAACTCTGTG 60
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 151 ATGGCTCACTATATACATTTCTCTGCATGGTTTGGTGGCTGCTTTCAGAACTCTGTG 210
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 61 TTAGCTGAAGTGGGGAAGTAAAGTCAAGTTGTCGTACTGCTCCGACAGATTTAGTTTC 120
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 211 TTAGCTGAAGTGGGGAAGTAAAGTCAAGTTGTCGTACTGCTCCGACAGATTTAGTTTC 270
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 121 ATCTTAGATGGCTCTTATAGTGTGGCCGACAAACCTTTGAAATAGTGAAGTGGCTT 180
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 271 ATCTTAGATGGCTCTTATAGTGTGGCCGACAAACCTTTGAAATAGTGAAGTGGCTT 330
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 181 GTCAATATCAGAAAACCTTTGACATAGGCGCGAAGTTTATTCAGTTGGAGTGGTTCAA 240
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 331 GTCAATATCAGAAAACCTTTGACATAGGCGCGAAGTTTATTCAGTTGGAGTGGTTCAA 390
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 241 TATAGTGAATACCTGCTCGAGATTCCTCTCGGAAGCTATGATTGAGGAGAACATTTG 300
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 391 TATAGTGAATACCTGCTCGAGATTCCTCTCGGAAGCTATGATTGAGGAGAACATTTG 450
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 301 ACGGACAGAGTGGATCCATCTCTACTTAGAGGAAACACAAAGACAGGAGGCCATC 360
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 451 ACGGACAGAGTGGATCCATCTCTACTTAGAGGAAACACAAAGACAGGAGGCCATC 510
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 361 CAGTTTGGCTCGATTA 377
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 511 CAGTTTGGCTCGATTA 527
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 6
 AI434106/c
 LOCUS
 DEFINITION
 t131b11.x1 NCI CGAP Gas4 Homo sapiens cDNA clone IMAGE:2132061 3'
 similar to TR:000339 000339 MATRILIN-2 ;, mRNA sequence.
 ACCESSION
 AI434106
 AI434106.1 GI:4293857
 EST.

SOURCE
 ORGANISM
 Homo sapiens
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 276)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished
 CONTACT: Robert Strausberg, Ph.D.
 Email: rgapbs@mail.nih.gov
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 893 Std Error: 0.00
 Seq primer: -40UP from Gibco.

FEATURES

Source

1..276
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2132061"
 /tissue_type="poorly differentiated adenocarcinoma with
 signet ring cell features"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Gas4"
 /note="Organ: stomach; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.69 kb. Life Technologies catalog #:
 '11543-011"
 82 a 61 c 39 g 94 t

BASE COUNT 82 a 61 c 39 g 94 t
 ORIGIN
 Query Match 9.6%; Score 276; DB 9; Length 276;
 Best Local Similarity 100.0%; Pred. No. 7.6e-129;
 Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 ACATAGTAGATAACATTATTGCTATTGGTCTGCTTCAGAAACAGAGATGCCGAAC 536
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 276 ACATAGTAGATAACATTATTGCTATTGGTGGTGGTTCAGAAACAGAGATGCCGAAC 217
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 537 TAGAGCTATTGCCAACAAAGCCCTTCGTCTACTTATGCTTTTATGTGGAAGACTATATTGC 596
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 216 TAGAGCTATTGCCAACAAAGCCCTTCGTCTACTTATGCTTTTATGTGGAAGACTATATTGC 157
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 597 AATATCCAAAATTAAGGGAAGTATGAAGCAGAAACCTTTGTGAGAACTCTGCTGCCAAC 656
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 156 AATATCCAAAATTAAGGGAAGTATGAAGCAGAAACCTTTGTGGAAGAACTCTGCTGCCAAC 97
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 657 ACGAATTCAGTGGCAGCTCGTGATGAAGGGGATTTGATATTCCTTTAGGTTTAGATGT 716
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 96 ACGAATTCAGTGGCAGCTCGTGATGAAGGGGATTTGATATTCCTTTAGGTTTAGATGT 37
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 717 AATAAAGAGTTAAGAAAGAAATACAGCTTTCACC 752
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 36 AATAAAGAGTTAAGAAAGAAATACAGCTTTCACC 1

RESULT 7
 N46880
 LOCUS
 DEFINITION
 N46880.1 Soares multiple sclerosis 2NBHMP Homo sapiens cDNA
 clone IMAGE:277704 5' similar to SW:CA1H MOUSE P39061 COLLAGEN
 ALPHA 1(XVII); CHAIN PRECURSOR. ;, mRNA sequence.
 ACCESSION
 N46880
 N46880.1 GI:1188C46
 EST.
 KEYWORDS
 Homo sapiens (human);

```

Db      188  GGAGAACAGGAGACATGGAAGGATGGATTAAATGGGTAGTCCCGGTTTCAAGGAGAA 247
Qy      1756  GCAGGATCCCTGCTGCTCCGGGCGAGATGGACACGGGGAGAGCTGGAATCCCCAGCA 1815
Db      248  GCAGGATCCCTGCTGCTCCGGGCGAGATGGACACGGGGAGAGCTGGAATCCCCAGCA 307
Qy      1816  TTTCCTGGAACCGAGGATTAATGGGCCAAAGAGGAGAAATTGGGCTCCAGGACAGCAA 1875
Db      308  TTTCCTGGAACCGAGGATTAATGGGCCAAAGAGGAGAAATTGGGCTCCAGGACAGCAA 367
Qy      1876  GGAAAAAAGAGAGCCCGAGGATCCCTGGTTTAAATGGGAAGCAATGGCTCACCAGGCCAG 1935
Db      368  GGAAAAAAGAGAGCCCGAGGATCCCTGGTTTAAATGGGAAGCAATGGCTCACCAGGCCAG 427
Qy      1936  CCTGGAACACCGGGATCTAAGGGAAGCAAAAGGTGAACCTGGAATTCAGGGATGCTGGG 1995
Db      428  CCTGGAACACCGGGATCTAAGGGAAGCAAAAGGTGAACCTGGAATTCAGGGATGCTGGG 487
Qy      1996  GCTTCTGGGCTCAAGGAGAACACAGGAGCAACGGGTTTCCCGAGAGAACAGGATACATG 2055
Db      488  GCTTCTGGGCTCAAGGAGAACACAGGAGCAACGGGTTTCCCGAGAGAACAGGATACATG 547
Qy      2056  GGTTTACCCGGGATTCAGGAAAGAAAGGGGACAAAGGAATCAAGGTGAAGAGGATTT 2115
Db      548  GGTTTACCCGGGATTCAGGAAAGAAAGGGGACAAAGGAATCAAGGTGAAGAGGATTT 607
Qy      2116  CAGGCTCAAAAGGGAGAGAAATGGAAGACAGGGAATTCAGGGCAACAGGGAATTCAGGC 2175
Db      608  CAGGCTCAAAAGGGAGAGAAATGGAAGACAGGGAATTCAGGGCAACAGGGAATTCAGGC 667
Qy      2176  CATCATGGTCAAAAGGAGAGAGAGGTGAAGAGGAGAACTGTGTTCAGAGTGCATTT 2235
Db      668  CATCATGGTCAAAAGGAGAGAGAGGTGAAGAGGAGAACTGTGTTCAGAGTGCATTT 727
Qy      2236  GGATCAAAAGGAGATCTGGG 2257
Db      728  GGATCAAAAGGAGATCTGGG 749

RESULT 4
BX431407 909 bp mRNA linear EST 15-MAY-2003
LOCUS BX431407 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE009YK07
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION BX431407
VERSION BX431407.1 GI:30773057
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li.W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3377.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAG0132B02_CS01162_1&cluster=3377.r.
Contact : Feng liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAG0132B02_CS01162_1.
Location/Qualifiers
1. 909
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE009YK07"

```

FEATURES

```

source
1. 909
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE009YK07"

```

```

/tissue type="PLACENTA"
/clore lib="Homo sapiens PLACENTA"
/Note=Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized.
BASE COUNT 251 a 201 c 205 g 249 t 3 others
ORIGIN

```

```

Query Match 15.23; Score 436; DB 13; Length 909;
Best Local Similarity 100.0%; Pred. No. 2.8e-210;
Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2430 AATTAGAAATTTGATCATTTGCTGTCCCAACATGGTCCCGGGTATTCCTGGGCCACC 2489
Db 35 AATTAGAAATTTGATCATTTGCTGTCCCAACATGGTCCCGGGTATTCCTGGGCCACC 94
Qy 2490 TGTTCGGATAGGCCAGGAGGTCCAGAGGATTTACCTGGTTTGGCAGGAAGAGATGGTGT 2549
Db 95 TGTTCGGATAGGCCAGGAGGTCCAGAGGATTTACCTGGTTTGGCAGGAAGAGATGGTGT 154
Qy 2550 TCCTGGATTTAGTGGTGTCTCTGACGTCCAGGTGTCCAGAGGATTTAAAGGCTTACCAGG 2609
Db 155 TCCTGGATTTAGTGGTGTCTCTGACGTCCAGGTGTCCAGAGGATTTAAAGGCTTACCAGG 214
Qy 2610 AAGAAATGGGAAAGAGGAGCCCAAGGGTTTGGGTATCTCTGGAGAACAGGTCTCTCTGG 2669
Db 215 AAGAAATGGGAAAGAGGAGCCCAAGGGTTTGGGTATCTCTGGAGAACAGGTCTCTCTGG 274
Qy 2670 TCCCGCAGGTCCAGAGGCCCTCTCTGGAATTAAGCAAGAGAGGTCTCCAGGAGACCAGG 2729
Db 275 TCCCGCAGGTCCAGAGGCCCTCTCTGGAATTAAGCAAGAGAGGTCTCCAGGAGACCAGG 334
Qy 2730 TCTCCTCGCAAGATGGAGACCATGGAACCTCGAATCCAGGCGCAACAGGCCCCCCC 2789
Db 335 TCTCCTCGCAAGATGGAGACCATGGAACCTCGAATCCAGGCGCAACAGGCCCCCCC 394
Qy 2790 AGGCATCTGGACCCCATCATTATTTTAGTGTAAATTCAGCAAGAGAGATCCGTTTCAAAA 2849
Db 395 AGGCATCTGGACCCCATCATTATTTTAGTGTAAATTCAGCAAGAGAGATCCGTTTCAAAA 454
Qy 2850 AGGACCAACTATTAG 2865
Db 455 AGGACCAACTATTAG 470

```

```

RESULT 5
CA395709 527 bp mRNA linear EST C6-NOV-2002
LOCUS CA395709
DEFINITION CS88a05.y1 Human Retinal pigment epithelium/choroid cDNA cs68a05
5', mRNA sequence.
(Un-normalized, unamplified): cs Homo sapiens cDNA clone cs68a05
ACCESSION CA395709
VERSION CA395709.1 GI:24731429
KEYWORDS EST
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 527)
REFERENCE 1 (bases 1 to 527)
AUTHORS Wischman,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behai,A.,
Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
Expressed sequence tag analysis of human RPE/choroid for the
NEBmark Project: Over 6000 non-redundant transcripts, novel genes
and splice variants
Mol. Vis. 8 (4), 205-220 (2002)
JOURNAL MEDLINE 22103460
PUBMED 12107410
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA

```

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: c9abbs-remail.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM13569 row: h column: 01
 High quality sequence stop: 693.
 Location/Qualifiers
 1. 929
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6182832"
 /sex="male"
 /tissue_type="dorsal root ganglia"
 /dev_stage="adult, 36 yr"
 /lab_host="DH10B"
 /clone_lib="Lupski dorsal root ganglion"
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site 1:
 NotI; Site 2: SalI; cDNA made by oligo-dT priming.
 Directionally cloned using the following adaptors:
 5'-TCGACCCACGGCTCCG-3' and
 5'-GACTAGTTCTAGATCCGAGCGCGCCCT(15)-3'. Size selected >
 1 kb for average insert length 1.7 kb. This is a primary
 library, non-amplified. Library constructed by Life
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
 College of Medicine) and is available through Life
 Technologies."

BASE COUNT 254 a 209 c 256 g 210 t

FEATURES

source

Query Match 25.1%; Score 720; DB 13; Length 929;

Best Local Similarity 100.0%; Pred.No. 0;

Matches 720; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2146 GGAATTCAGGCGCAACAGGGAATTCAGGCCATCATGTGTCAGAAAGAGAGAGAGGTGAA 2205

DB 1 GGAATTCAGGCGCAACAGGGAATTCAGGCCATCATGTGTCAGAAAGAGAGAGAGGTGAA 60

QY 2206 AAGGAGAACCTGGTGTCCGAGGTGCCATTGGATCAAAAGAGAGATCTGGGTGGATGGC 2265

DB 61 AAGGAGAACCTGGTGTCCGAGGTGCCATTGGATCAAAAGAGAGATCTGGGTGGATGGC 120

QY 2266 TTGATGGGCGCCGACAGGTCCCTAAGGGCAACCTGGGGATCCAGGTCCCTCAGGGACCCCA 2325

DB 121 TTGATGGGCGCCGACAGGTCCCTAAGGGCAACCTGGGGATCCAGGTCCCTCAGGGACCCCA 180

QY 2326 GGTGGATGGGAAGCCCGGAGAGAGATTTTCAGAACAAATTTATTCGACAAAGTTTCGACA 2385

DB 181 GGTGGATGGGAAGCCCGGAGAGAGATTTTCAGAACAAATTTATTCGACAAAGTTTCGACA 240

QY 2386 GATGTAATAAGACCCAGCTACCACTCTTACTTCAGACTGGAAGATTTAGAAATTTGAT 2445

DB 241 GATGTAATAAGACCCAGCTACCACTCTTACTTCAGACTGGAAGATTTAGAAATTTGAT 300

QY 2446 CATTCGCTGTCCCAACATGGGTCCCGGGTATTCCTGGGCCACCTGGTCCGATAGGCCCA 2505

DB 301 CATTCGCTGTCCCAACATGGGTCCCGGGTATTCCTGGGCCACCTGGTCCGATAGGCCCA 360

QY 2506 GAGGTGCCAGAGATTTACCTGGTTTGCAGAGAGATGTGTTCCTGGATTTAGTGGT 2565

DB 361 GAGGTGCCAGAGATTTACCTGGTTTGCAGAGAGATGTGTTCCTGGATTTAGTGGT 420

QY 2566 CTCCTCGACGTCCAGGTGTTCAGAGATTTAAGAGCCCTACAGAGAGATTTGGGAAAAA 2625

DB 421 CTCCTCGACGTCCAGGTGTTCAGAGATTTAAGAGCCCTACAGAGAGATTTGGGAAAAA 480

QY 2626 GGCAGCCAGGGTTTGGGTATCTCTGAGAACAGGTCTCTCTGGTCCCGCCAGGTCCAGAG 2695

DB 481 GGGAGCCAAAGGTTTGGGTATCTCTGAGAACAGGTCTCTCTGGTCCCGCCAGGTCCAGAG 540

QY 2686 GGGCTCTCTGGATAGCAAAAGAGTCTCTCCAGAGACCCAGGTCTCTCTGGCAAAGAT 2745

DB 541 GGGCTCTCTGGATAGCAAAAGAGTCTCTCCAGAGACCCAGGTCTCTCTGGCAAAGAT 600

QY 2746 GGAGACCATGGAAAAACCTGGAAATCCAAAGGGCAACAGGCCCCCAGGATCTCGACCCCA 2805

DB 601 GGAGACCATGGAAACCTGGAAATCCAAAGGGCAACAGGCCCCCAGGATCTCGACCCCA 660

QY 2806 TCATATGTTTAAAGTGTATTTCCAGAGAGATTCGGTTTCAGAAAAAGCCAAACTATTAG 2865

DB 661 TCATATGTTTAAAGTGTATTTCCAGAGAGATTCGGTTTCAGAAAAAGCCAAACTATTAG 720

RESULT 3

EX458795

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

EX458795 Homo sapiens PLACENTA Homo sapiens cDNA clone CSDE009YK07
 5-PRIME, mRNA sequence.
 EX458795
 EST.
 Homo sapiens (human)
 Homo sapiens
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1201)
 Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished
 JOURNAL
 COMMENT

Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 3377.r For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CSDE009AF040P1&cluster=3377.r. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1603
 Faraday Avenue Genoscope sequence ID : CSDE009AF040P1.

FEATURES
 source
 1. 1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CSDE009YK07"
 /tissue_type="PLACENTA"
 /clone_lib="Homo sapiens PLACENTA"
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-cligc(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

BASE COUNT 350 a 223 c 387 g 212 t 29 others

ORIGIN

Query Match 23.8%; Score 682; DB 13; Length 1201;

Best Local Similarity 100.0%; Pred.No. 0;

Matches 682; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1576 CCAGATCAAAAGGTTGAAATGGTGCACAAAGAGACAAAGGATCACCTGGATTTATGGC 1635

DB 68 CCAGATCAAAAGGTTGAAATGGTGCACAAAGAGACAAAGGATCACCTGGATTTATGGC 127

QY 1636 AAAAAAGGTTGAAAGGTTGAAAGGGAATGTGGTCCCTGGCTCCCTGGACTGCT 1695

DB 128 AAAAAAGGTTGAAAGGTTGAAAGGGAATGTGGTCCCTGGCTCCCTGGACTGCT 187

QY 1696 GAGAGACCAAGGATGAGATGAGATTAATGGGTAGTCCCGTTTCAGAGGAA 1755

QY	961	TTATTTTACAACAAACCAGCGTAATTAAATGGCTCACAAGTGTTTACCTTTGCTAACCCCTCAA	1020
Db	1148	TTATTTTACAACAAACCAGCGTAATTAAATGGCTCACAAGTGTTTACCTTTGCTAACCCCTCAA	1207
QY	1021	GTTAAGACGTTGTGTGATGAAGCGTCGCACCAAAATTCGTCTCTTAGTAAACAGAACAGAT	1083
Db	1208	GTTAAGACGTTGTGTGATGAAGCGTCGCACCAAAATTCGTCTCTTAGTAAACAGAACAGAT	1267
QY	1081	GTGACTTTGTATATTGTATGATGACCAACAAATTTGAAAACAAAGCCCTTACATCCAGTTTTAGGG	1140
Db	1268	GTGACTTTGTATATTGTATGATGACCAACAAATTTGAAAACAAAGCCCTTACATCCAGTTTTAGGG	1327
QY	1141	ATCTTGATCAATGGGCAAAACCCAAATTTGGAATAATTTCTGGAAGAAGAAATGTTTCAG	1200
Db	1328	ATCTTGATCAATGGGCAAAACCCAAATTTGGAATAATTTCTGGAAGAAGAAATGTTTCAG	1387
QY	1201	TTTGTATGTCMAAAGTTTGCGAATCTACTGTGACCCAGAAACAGAACCAACCGGAGACAGCA	1260
Db	1388	TTTGTATGTCMAAAGTTTGCGAATCTACTGTGACCCAGAAACAGAACCAACCGGAGACAGCA	1447
QY	1261	TGTGAGATTCCTGGATTT-----TGCCTTAATGGTCCCAGTGTATAGTTCAACT	1311
Db	1448	TGTGAGATTCCTGGATTTAAATGGAGATGGCTTAAATGGTCCCAGTGTATAGTTCAACT	1507
QY	1312	CCAGCTCCCCGTATTTGTGCTCCCGGAAAAACAGAGCTTCAAGGCCCCCAAGGTGACCCCT	1371
Db	1508	CCAGCTCCCTGTATTTGTCTCCCGGAAAAACAGAGCTTCAAGGCCCCCAAGGTGACCCCT	1567
QY	1372	GGACTGCTCGGAAACCTCGCTACCTTGGACACACCTGGTCAAGATGTAAGCCTGGATAT	1431
Db	1568	GGACTGCTCGGAAACCTCGCTACCTTGGACACACCTGGTCAAGATG---GGATAT	1619
QY	1432	CAGGGAATTGCAGGAGCACACAGGTGTTCCAGGATCTCCAGGAATACAAAGCAGCTCCAGGA	1491
Db	1620	CAGGGAATTGCAGGAGCACACAGGTGTTCCAGGATCTCCAGGAATACAAAGCAGCTCCAGGA	1679
QY	1492	CTACAGGTTTACAAAGGAGAACACAGGCGAGATGGTGACAAAGGATGATCTGGAATTCCT	1551
Db	1680	CTACAGGTTTACAAAGGAGAACACAGGCGAGATGGTGACAAAGGATGATCTGGAATTCCT	1739
QY	1552	GGTTTTCCTGGCTTCATGGCATCCAGGATCAAAAGGTCMAATGGGTGCCAAGNGAC	1611
Db	1740	GGTTTTCCTGGCTTCATGGCATCCAGGATCAAAAGGTCMAATGGGTGCCAAGNGAGAC	1799
QY	1612	AAAGSATCACCTCGATTTTATGGCAAAAAGGTCGAAAAGGTGAAAAGGGGAATGCCGGC	1671
Db	1800	AAAGSATCACCTCGATTTTATGGCAAAAAGGTCGAAAAGGTGAAAAGGGGAATGCTGGC	1859
QY	1672	TTCCCTGGCCCTCCCTGGACCTGCTGGAGAACCCAGGAGACATGGAAGAAGTATGATTAATG	1731
Db	1860	TTCCCTGGCCCTCCCTGGACCTGCTGGAGAACCCAGGAGACATGGAAGAAGTATGATTAATG	1919
QY	1732	GGTAGTCCCGCTTTCAAGGGAGAACAGGATCCCTGGTGCTCCGGGGCAGGATGGAACA	1791
Db	1920	GGTAGTCCCGCTTTCAAGGGAGAACAGGATCCCTGGTGCTCCGGGGCAGGATGGBACA	1979
QY	1792	CGGGAGAGCCTGGAATCCCHAGNATTTCTGGAAACCGAGGATTAATGGGCGCAAAAGGGA	1851
Db	1980	CGGGAGAGCCTGGAATCCCHAGNATTTCTGGAAACCGAGGATTAATGGGCGCAAAAGGGA	2039
QY	1852	GAAATTTGGGCTCCAGGACACCAAGGAATAAAGAGCCCCAGGATGCTCTGGTTTAATG	1911
Db	2040	GAAATTTGGGCTCCAGGACACCAAGGAATAAAGAGCCCCAGGATGCTCTGGTTTAATG	2099
QY	1912	GGAGCAATTGCTCACCAAGCCAGCTGGAAACACCGGATCTAAGGGAAGCAAGGTGAA	1971
Db	2100	GGAGCAATTGCTCACCAAGCCAGCTGGAAACACCGGATCTAAGGGAAGCAAGGTGAA	2159
QY	1972	CCTGGAATTCAGGGATGCCTTGGGGCTTCTGGGCTCAAGGAGAACCCAGGAGCAACGGGT	2031
Db	2160	CCTGGAATTCAGGGATGCCTTGGGGCTTCTGGGCTCAAGGAGAACCCAGGAGCAACGGGT	2219

QY	2032	TC	CC	AGGAGAA	CC	AGGATAC	TG	GGT	TTT	AC	CGG	AT	TT	AC	CGG	AT	TT	CA	AG	AAAA	AAAA	AGG	GGG	CAAA	2091							
DB	2220	TC	CC	AGGAGAA	CC	AGGATAC	TG	GGT	TTT	AC	CGG	AT	TT	AC	CGG	AT	TT	CA	AG	AAAA	AAAA	AGG	GGG	CAAA	2279							
QY	2092	GGA	AA	TC	CA	AGT	GA	AA	AGT	TAT	TC	AGG	GT	CA	AA	AGG	CA	AA	AA	TC	GA	AG	CA	AGG	AA	2151						
DB	2280	GGA	AA	TC	CA	AGT	GA	AA	AGT	TAT	TC	AGG	GT	CA	AA	AGG	CA	AA	AA	TC	GA	AG	CA	AGG	AA	2339						
QY	2152	CC	AGG	CA	AA	CAG	GA	AT	TC	AA	G	CC	AT	CA	TG	GC	AA	AA	AG	AG	AG	AG	AG	AG	GA	2211						
DB	2340	CC	AGG	CA	AA	CAG	GA	AT	TC	AA	G	CC	AT	CA	TG	GC	AA	AA	AG	AG	AG	AG	AG	AG	GA	2399						
QY	2212	GA	AC	CT	CG	TG	CC	AG	GT	GC	CA	TT	TG	GA	T	CA	AA	AG	GA	AT	TC	CGG	T	GG	AT	GC	2271					
DB	2400	GA	AC	CT	CG	TG	CC	AG	GT	GC	CA	TT	TG	GA	T	CA	AA	AG	GA	AT	TC	CGG	T	GG	AT	GC	2459					
QY	2272	GG	CC	CC	CA	AG	GT	CC	TC	AA	G	GG	CA	AC	TC	CG	AG	T	CC	AG	GT	CC	TC	CG	AG	GT	2331					
DB	2460	GG	CC	CC	CA	AG	GT	CC	TC	AA	G	GG	CA	AC	TC	CG	AG	T	CC	AG	GT	CC	TC	CG	AG	GT	2519					
QY	2332	GAT	GG	GA	AG	CC	CG	GA	CAG	AT	TT	TC	CA	GA	CA	AT	TT	AT	TC	GA	CA	AT	TT	TC	GA	CAG	AT	2391				
DB	2520	GAT	GG	GA	AG	CC	CG	GA	CAG	AT	TT	TC	CA	GA	CA	AT	TT	AT	TC	GA	CA	AT	TT	TC	GA	CAG	AT	2579				
QY	2392	ATA	AG	AC	CC	CAG	CT	TAC	CT	TACT	TC	CAG	AT	CG	AA	GA	AT	TAG	AA	AT	TC	GA	TC	AT	TC	GC	2451					
DB	2558	ATA	AG	AC	CC	CAG	CT	TAC	CT	TACT	TC	CAG	AT	CG	AA	GA	AT	TAG	AA	AT	TC	GA	TC	AT	TC	GC	2539					
QY	2452	CT	GT	CC	CA	AT	GC	CT	CC	CG	GT	AT	TC	TC	GG	CC	AC	TG	TC	CG	AT	AG	CC	CC	AG	AG	GT	2511				
DB	2640	CT	GT	CC	CA	AT	GC	CT	CC	CG	GT	AT	TC	TC	GG	CC	AC	TG	TC	CG	AT	AG	CC	CC	AG	AG	GT	2699				
QY	2512	CC	CA	GA	GA	T	AC	CT	GG	TT	TC	CC	AG	GA	GA	T	GG	T	TC	CT	CG	AT	TC	CG	AT	TC	GC	2571				
DB	2720	CC	CA	GA	GA	T	AC	CT	GG	TT	TC	CC	AG	GA	GA	T	GG	T	TC	CT	CG	AT	TC	CG	AT	TC	GC	2759				
QY	2572	GG	AG	TC	CA	AG	GT	TC	CA	AG	GA	T	TA	AA	AG	GC	T	AC	CA	GA	GA	AA	T	GG	GA	AA	GG	AG	GC	2631		
DB	2760	GG	AG	TC	CA	AG	GT	TC	CA	AG	GA	T	TA	AA	AG	GC	T	AC	CA	GA	GA	AA	T	GG	GA	AA	GG	AG	GC	2631		
QY	2632	CA	AG	GT	TC	CG	AT	TC	CG	GA	CA	AG	GT	TC	CT	CG	TC	CC	CC	CC	AG	GT	TC	CG	AG	GT	TC	GC	2691			
DB	2820	CA	AG	GT	TC	CG	AT	TC	CG	GA	CA	AG	GT	TC	CT	CG	TC	CC	CC	CC	AG	GT	TC	CG	AG	GT	TC	GC	2879			
QY	2692	CCT	GG	AA	TA	AG	CA	AG	GT	TC	CT	CC	AG	GA	CA	CC	AG	GT	TC	CT	CG	TC	GG	CA	AA	AG	AT	GG	AG	CA	2751	
DB	2880	CCT	GG	AA	TA	AG	CA	AG	GT	TC	CT	CC	AG	GA	CA	CC	AG	GT	TC	CT	CG	TC	GG	CA	AA	AG	AT	GG	AG	CA	2939	
QY	2752	CAT	GG	AA	AA	CT	CG	AA	T	CC	AA	G	GC	CA	AG	GG	CC	CC	CC	CC	AG	GC	AT	CT	CG	AC	CC	AT	CAC	TA	2811	
DB	2940	CAT	GG	AA	AA	CT	CG	AA	T	CC	AA	G	GC	CA	AG	GG	CC	CC	CC	CC	AG	GC	AT	CT	CG	AC	CC	AT	CAC	TA	2999	
QY	2812	TG	TT	T	AG	T	GA	T	T	GC	C	A	A	G	A	G	A	T	T	C	C	G	T	T	C	A	G	A	A	A	A	2865
DB	3000	TG	TT	T	AG	T	GA	T	T	GC	C	A	A	G	A	G	A	T	T	C	C	G	T	T	C	A	G	A	A	A	A	3053
RESULT 2																																
BQ893743																																
LOCUS																																
DEFINITION		AGENCOURT 8232060 Lupski dorsal root ganglion Homo sapiens CDNA clone IMAGE:6182832 5', mRNA sequence.																														
ACCESSION		BQ893743																														
VERSION		BQ893743.1 GI:22285757																														
KEYWORDS		ES.																														
SOURCE		Homo sapiens (human)																														
ORGANISM		Homo sapiens																														
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.																														
AUTHORS		1. (bases 1 to 929)																														
TITLE		NH-MGC http://mgc.nci.nih.gov/																														
JOURNAL		National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished																														

```

80 20 0.7 186 9 AI906933
79 20 0.7 246 14 CB241974
80 20 0.7 290 10 BF652344
81 20 0.7 296 9 AW685176
82 20 0.7 296 13 BQ557911
83 20 0.7 297 28 B43380
84 20 0.7 323 10 BG583481
85 20 0.7 326 9 AI301744
86 20 0.7 345 28 AZ056334
87 20 0.7 347 28 AQ167717
88 20 0.7 351 28 BH337770
89 20 0.7 353 9 AW356726
90 20 0.7 369 28 AZ039808
91 20 0.7 371 14 CA909116
92 20 0.7 376 29 TA122H03Q
93 20 0.7 389 10 BF655157
94 20 0.7 397 9 AA115401
95 20 0.7 406 10 BE284841
96 20 0.7 415 14 C82609
97 20 0.7 415 14 C83465
98 20 0.7 429 14 R23059
99 20 0.7 430 9 AV733646
100 20 0.7 438 9 AV594520

```

ALIGNMENTS

```

RESULT 1
BC045597 4145 bp mRNA linear HTC 04-MAR-2003
LOCUS BC045597 Homo sapiens, clone IMAGE:4814199, mRNA.
DEFINITION BC045597.1 GI:28374134
ACCESSION BC045597.1
VERSION 1
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Strausberg,R.
Direct Submission
Submitted (31-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
CDNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 32 Row: K Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18780272
This clone has the following problem: retained intron.

FEATURES

```

source
1..4145
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4814199"
/tissue_type="Brain, hippocampus"

```

```

/clone_lib="NIH_MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
BASE COUNT 1275 a 787 c 1017 g 1066 t
ORIGIN
Query Match 54.6%; Score 1564; DB 11; Length 4145;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2854; Conservative 0; Mismatches 3; Indels 17; Gaps 2;
QY 1 ATGGCTGACTATATACATTTCTGCTGATGGTTGGTGGTGGTCTTCCAGAACTCTGTG 60
DB 188 AAGGCTCACTATATTACATTTCTGCTGATGGTTGGTGGTGGTCTTCCAGAACTCTGTG 247
QY 61 TTAGCTGAAGATGGGGAAGTCAAGTGTCTGCTACTGCTCCGACAGAGATTATTTTC 120
DB 248 TTAGCTGAAGATGGGGAAGTCAAGTGTCTGCTACTGCTCCGACAGAGATTATTTTC 307
QY 121 ACTTTAGATGCTCTTTATAGTGTGGCCGAGAACTTTGAAATAGTGAAGAGTGGCTT 190
DB 306 ATCTTAGATGCTCTTTATAGTGTGGCCGAGAACTTTGAAATAGTGAAGAGTGGCTT 367
QY 181 GTCATATACAAAAAATTTTGACATAGGCGCGAGTTTATTCAGTTGGAGTGGTCAA 240
DB 368 GTCATATACAAAAAATTTTGACATAGGCGCGAGTTTATTCAGTTGGAGTGGTCAA 427
QY 241 TATAGTACTACCTGTGCTGAGATTTCTCTCGAAGCTATATGATTCAGGAGAACTTTG 300
DB 428 TATAGTACTACCTGTGCTGAGATTTCTCTCGAAGCTATATGATTCAGGAGAACTTTG 487
QY 301 ACGGAGAGTGGATTCATCTACTTAGGAGGAAACACAAAGACAGAGGAGGCGCATC 360
DB 488 ACGGAGAGTGGATTCATCTACTTAGGAGGAAACACAAAGACAGAGGAGGCGCATC 547
QY 361 CAGTTTGGCTCGATTACCTTTTGGCAAGCTCTCAGATTCTGACTTAAGATAGCAGTG 420
DB 548 CAGTTTGGCTCGATTACCTTTTGGCAAGCTCTCAGATTCTGACTTAAGATAGCAGTG 607
QY 421 GTACTTACCGATGGCAATCCCAAGATGACGTCAAGGATGACGTCAAGCAGCAGAGAT 480
DB 608 GTACTTACCGATGGCAATCCCAAGATGACGTCAAGGATGACGTCAAGCAGCAGAGAT 667
QY 481 AGTAAGATAAATTTATTTGCTATTGTTGGTTCAGAAACAGAGAGTCCCGAATCTAGA 540
DB 668 AGTAAGATAAATTTATTTGCTATTGTTGGTTCAGAAACAGAGAGTCCCGAATCTAGA 727
QY 541 GCTATTGCCAACAGAGCTTCTGCTACTTATCTGTTTATGTGGAAGACTATATGCAATA 600
DB 728 GCTATTGCCAACAGAGCTTCTGCTACTTATCTGTTTATGTGGAAGACTATATGCAATA 787
QY 601 TCCAAAATTAAGGGAAGTGAATGAAGCAGAACTTTTGTGAAGAACTCTGTGTCACACGA 660
DB 788 TCCAAAATTAAGGGAAGTGAATGAAGCAGAACTTTTGTGAAGAACTCTGTGTCACACGA 847
QY 661 ATTCCAGTGGCAGCTCGTGATGAAGGGGATTTGATATTCTTTAGTTAGATGAAT 720
DB 948 ATTCAGTGGCAGCTCGTGATGAAGGGGATTTGATATTCTTTAGTTAGATGAAT 907
QY 721 AAAAGGTTAAGAAAAGAAATACAGCTTTTCCAAAAAAGATATAAGATTAACA 780
DB 908 AAAAGGTTAAGAAAAGAAATACAGCTTTTCCAAAAAAGATATAAGATTAACA 967
QY 781 TCAAAGTTGATTTATCAGAACTCAGAACTCAGAACTTTTCCAGAGGTTCTTCTCCATCA 840
DB 968 TCAAAGTTGATTTATCAGAACTCAGAACTCAGAACTTTTCCAGAGGTTCTTCTCCATCA 1027
QY 841 TATGATTTTGTGCTACTCAAGATTTTAAAGTCAAGAAAATTTGGGATTTATGAGAGATA 900
DB 1028 TATGATTTTGTGCTACTCAAGATTTTAAAGTCAAGAAAATTTGGGATTTATGAGAGATA 1087
QY 901 TTAATTTTGAAGAGGCGCAAAATAGCAGTTTACCTTAATGCTGTGGACAAATCTTA 960
DB 1086 TTAATTTTGAAGAGGCGCAAAATAGCAGTTTACCTTAATGCTGTGGACAAATCTTA 1147

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 22:19:29 ; Search time 5306 Seconds

(without alignments)
13123.318 Million cell updates/sec

Title: US-09-996-611B-5

Perfect score: 2865

Sequence: 1 atggtctactattacatt.....gaaaaggacaaactattag 2865

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

EST:*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_htc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gas_hum.*

18: em_gas_inv.*

19: em_gas_pln.*

20: em_gas_vrt.*

21: em_gas_fun.*

22: em_gas_mam.*

23: em_gas_mus.*

24: em_gas_pro.*

25: em_gas_rod.*

26: em_gas_pbg.*

27: em_gas_vrl.*

28: gb_gasl.*

29: gb_gasl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query	ID	Description
1	1564	54.6	4145	11	BC045597 Homo sapi
2	720	25.1	929	13	BQ93743 AGNCOURT
3	682	23.8	1201	13	BX458795 BX458795
4	436	15.2	909	13	BX431407 BX431407

CA395709	CS68a05.Y	527	14	CA395709
AI434106	ti31b11.x	276	9	AI434106
N46880	YY58a01.r1	245	7	N46880
AA227783	zr57h04.r	233	8	AA227783
AI127272	qb70c09.x	218	9	AI127272
N84635	JC337F.Huma	165	5	N84635
AW962485	EST374558	164	10	AW962485
BG899698	602681559	163	11	BG899698
T70851	yd15e11.r1	161	12	T70851
AA095402	14401.seq	156	13	AA095402
N86630	J9414F.Huma	129	14	N86630
CB852087	UI-CF-PNO	112	15	CB852087
CA329319	EST32913	106	16	CA329319
N85186	J2812F.Huma	104	17	N85186
CA397287	cs83c07.Y	85	18	CA397287
R58412	G3577.Fetal	74	19	R58412
BE701852	CM1--NN100	66	20	BE701852
CB241974	UI-CF-PNO	64	21	CB241974
AA247907	sl114.seq	58	22	AA247907
CB268816	1007722.H	57	23	CB268816
AL602341	DXFZP686B	56	24	AL602341
BE236040	143702.MA	45	25	BE236040
R77744	Y180902.sl	44	26	R77744
AQ356449	CITBI-EI-	42	27	AQ356449
EX391778	EX391778	38	28	EX391778
AQ836646	HS-5427.B	36	29	AQ836646
BE899929	18C456.MA	35	30	BE899929
AQ514534	HS-5081.B	34	31	AQ514534
CC866742	CSU-K33-	33	32	CC866742
AV595980	AV595980	32	33	AV595980
CB801379	AMGNNUC-M	22	34	CB801379
AV597967	AV597967	22	35	AV597967
AV614172	AV614172	22	36	AV614172
BH882289	hw36d12.S	22	37	BH882289
AV599587	AV599587	22	38	AV599587
AV592493	AV592493	22	39	AV592493
BH882288	hw36d12.b	22	40	BH882288
AL200803	Tetradon	22	41	AL200803
SE437463	SFR004.AC	22	42	SE437463
CC435130	PUHQJ049TD	22	43	CC435130
CB392620	OSTR103CS	22	44	CB392620
BB599402	BB599402	22	45	BB599402
BB605733	BB605733	22	46	BB605733
EG518121	947067F03	22	47	EG518121
BY027337	BY027337	22	48	BY027337
BY001738	BY001738	22	49	BY001738
AQ612958	HS-5129.A	22	50	AQ612958
CB786369	AMGNNUC-N	22	51	CB786369
AA438085	vd22b07.S	22	52	AA438085
CD552725	E0348B10-	22	53	CD552725
BF211912	601813084	22	54	BF211912
CD390380	Gm_CK0596	22	55	CD390380
BU722040	SJYADA07	22	56	BU722040
CA685857	wlm96.pk0	22	57	CA685857
AI711999	614002F05	22	58	AI711999
BZ806868	PUGX072D	22	59	BZ806868
BH927741	odh24c03.	22	60	BH927741
BY735103	BY735103	22	61	BY735103
AQ262588	CITBI-EI-	22	62	AQ262588
CD351288	UI-M-GIO-	22	63	CD351288
BZ496469	BONMM94TF	22	64	BZ496469
BQ781082	UI-R-FF0-	22	65	BQ781082
BH45101	BOGEH42PR	22	66	BH45101
BG837006	Zm08_08f1	22	67	BG837006
BI646459	603279039	22	68	BI646459
AL258708	Tetradon	22	69	AL258708
AK085362	Mus_muscu	22	70	AK085362
AZ098728	RPCT-23-1	22	71	AZ098728
BE682427	150462.MA	22	72	BE682427

```

; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5386
; LENGTH: 3609
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-5386

Query Match      0.7%  Score 19;  DB 14;  Length 3609;
Best Local Similarity 100.0%;  Pred.No. 77;
Matches 19;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY 1570 GGCATGCCAGGATCAAGG 1589
    |||||||||||||||
Db 3026 GGCATGCCAGGATCAAGG 3008

Search completed: October 3, 2003, 05:18:13
Job time : 2736 secs

```


APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 111629
; LENGTH: 2652
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-111629

Query Match 0.7%; Score 19; DB 13; Length 2652;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1989 GCCTGGGGCTTCTGGGCTC 2007
|||||
DB 1748 GCCTGGGGCTTCTGGGCTC 1730

RESULT 43
US-10-027-632-111630/c
; Sequence 111630, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 111630
; LENGTH: 2652
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-111630

Query Match 0.7%; Score 19; DB 13; Length 2652;
Best Local Similarity 100.0%; Pred. No. 76;
Matches, 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1989 GCCTGGGGCTTCTGGGCTC 2007
|||||
DB 1748 GCCTGGGGCTTCTGGGCTC 1730

RESULT 44
US-10-128-714-386/c
; Sequence 386, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Weng
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128.714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 386
; LENGTH: 3608
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-386

Query Match 0.7%; Score 19; DB 14; Length 3608;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1570 GCATGCCAGGATCAAGG 1588
|||||
DB 3026 GCATGCCAGGATCAAGG 3008

RESULT 45
US-10-128-714-5386/c
; Sequence 5386, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Weng
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10128.714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362

TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-3872

Query Match 0.7%; Score 19; DB 10; Length 1363;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1101 CCAACAAATTGAAACAAAG 1119
|||||
DB 720 CCAACAAATTGAAACAAAG 702

RESULT 39
US-10-240-965-152/c
Sequence 152, Application US/10240965
Publication No. US20030165924A1
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: SHIFFMAN, Dov
APPLICANT: SOMOGYI, Roland
APPLICANT: LAWN, Richard M.
APPLICANT: SEILHAMER, Jeffrey J.
APPLICANT: PORTER, Gordon J.
APPLICANT: MIKITA, Thomas
APPLICANT: TAI, Julie
TITLE OF INVENTION: GENES EXPRESSED IN FOAM CELL DIFFERENTIATION
FILE REFERENCE: PA-0025 PCT
CURRENT APPLICATION NUMBER: US/10/240,965
CURRENT FILING DATE: 2002-10-04
PRIOR APPLICATION NUMBER: 60/195,106
PRIOR FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 276
SOFTWARE: PERL Program
SEQ ID NO 152
LENGTH: 2341
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030165924A1 245099.8
US-10-240-965-152

Query Match 0.7%; Score 19; DB 12; Length 2341;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2025 AACGGGTTCCCGAGGAGAA 2043
|||||
DB 104 AACGGGTTCCCGAGGAGAA 86

RESULT 40
US-10-125-237-52/c
Sequence 52, Application US/10125237
Publication No. US20030022329A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Zhou, Ping
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyan
APPLICANT: Zhao, Qing A.
APPLICANT: Xue, Aidong J.
APPLICANT: Zhang, Jie
APPLICANT: Wehrman, Tom
APPLICANT: Wang, Jian-Rui
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. US20030022329A1el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 791CIP2ADiv
CURRENT APPLICATION NUMBER: US/10/125,237
CURRENT FILING DATE: 2002-04-17

PRIOR APPLICATION NUMBER: 09/668,317
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: 09/552,929
PRIOR FILING DATE: 2000-04-18
NUMBER OF SEQ ID NOS: 91
SOFTWARE: P₁_FL_genes Version 2.0
SEQ ID NO 52
LENGTH: 2346
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (86) .. (1564)
US-10-125-237-52

Query Match 0.7%; Score 19; DB 14; Length 2346;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2025 AACGGGTTCCCGAGGAGAA 2043
|||||
DB 111 AACGGGTTCCCGAGGAGAA 93

RESULT 41
US-10-105-891-52/c
Sequence 52, Application US/10105891
Publication No. US20030073099A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Zhou, Ping
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyan
APPLICANT: Zhao, Qing A.
APPLICANT: Xue, Aidong J.
APPLICANT: Zhang, Jie
APPLICANT: Wehrman, Tom
APPLICANT: Wang, Jian-Rui
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. US20030073099A1el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 791CIP2A
CURRENT APPLICATION NUMBER: US/10/105,891
CURRENT FILING DATE: 2002-03-25
PRIOR APPLICATION NUMBER: 09/668,317
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: 09/552,929
PRIOR FILING DATE: 2000-04-18
NUMBER OF SEQ ID NOS: 91
SOFTWARE: P₁_FL_genes Version 2.0
SEQ ID NO 52
LENGTH: 2346
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (86) .. (1564)
US-10-105-891-52

Query Match 0.7%; Score 19; DB 14; Length 2346;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2025 AACGGGTTCCCGAGGAGAA 2043
|||||
DB 111 AACGGGTTCCCGAGGAGAA 93

RESULT 42
US-10-027-632-111629/c
Sequence 111629, Application US/10027632
GENERAL INFORMATION:

APPLICANT: Smith, Victoria
 ADDRESS: 201, Collyer Crescent
 CITY: Melbourne
 STATE: Victoria
 POSTAL CODE: 3000
 COUNTRY: Australia
 TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

```
; APPLICANT: Dougan, Gordon
; APPLICANT: Wilson, Rebecca Kerry
; TITLE OF INVENTION: Genes and Proteins, and Their Use
; FILE REFERENCE: GJE-70C1
; CURRENT APPLICATION NUMBER: US/10/096,162
; CURRENT FILING DATE: 2002-11-10
; PRIOR APPLICATION NUMBER: GB 0105922.9
; PRIOR FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: group B streptococcus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1146)
; OTHER INFORMATION:
; US-10-096-162-1
```

```
Query Match 0.7%; Score 19; DB 14; Length 1146;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2606 CAGGAGAAATGGGAAAA 2624
      |||||||
Db 556 CAGGAGAAATGGGAAAA 574
```

```
RESULT 32
US-10-027-632-117674
; Sequence 117674, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117674
; LENGTH: 1188
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-117674
```

```
Query Match 0.7%; Score 19; DB 13; Length 1188;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2297 CTGGGGATCCAGGTCCTCA 2315
      |||||||
Db 595 CTGGGGATCCAGGTCCTCA 613
```

```
RESULT 33
US-10-027-632-117675
; Sequence 117675, Application US/10027632
```

```
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117675
; LENGTH: 1188
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-117675
```

```
Query Match 0.7%; Score 19; DB 13; Length 1188;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2297 CTGGGGATCCAGGTCCTCA 2315
      |||||||
Db 595 CTGGGGATCCAGGTCCTCA 613
```

```
RESULT 34
US-10-027-632-117676
; Sequence 117676, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117676
; LENGTH: 1188
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-117676
```

```
Query Match 0.7%; Score 19; DB 13; Length 1188;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


LENGTH: 378361
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(378361)
OTHER INFORMATION: n = A,T,C or G
US-09-901-136-3

Query Match 0.7%; Score 20; DB 11; Length 378361;
Best Local Similarity 100.0%; Pred.No. 26;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2123 AAAAGGGAGAAATGGGA 2142
DB 10250 AAAAGGGAGAAATGGGA 10231

RESULT 26
US-09-263-959-1
Sequence 1, Application US/09263959
Patent No. US20020150891A1
GENERAL INFORMATION:
APPLICANT: Hood, Leroy E.
APPLICANT: Rowen, Lee
TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH UTI
NUMBER OF SEQUENCES: 1279
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263,959
FILING DATE: 05-MAR-1999
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Mcmasters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 920010.426C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 684973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-263-959-1

Query Match 0.7%; Score 20; DB 10; Length 684973;
Best Local Similarity 100.0%; Pred.No. 27;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1842 CCAAAGGGAGAAATGGGC 1861
DB 181777 CCAAAGGGAGAAATGGGC 181796

RESULT 27
US-09-783-590-8616/c
Sequence 8616, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:

APPLICANT: Dillon, Patrick J.
APPLICANT: Haseltine, William A.
APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products i6.2
FILE REFERENCE: PO-16.2C1
CURRENT APPLICATION NUMBER: US/09/783,590
CURRENT FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 08/420,856
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 8616
LENGTH: 381
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (27)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (37)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (38)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (58)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (72)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (155)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (178)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (210)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (222)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (257)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (259)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (313)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (330)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (340)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (355)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (361)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (365)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (368)

```
;
;
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 621:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 621:
US-09-070-927A-621

Query Match 0.7%; Score 20; DB 10; Length 1635;
Best Local Similarity 100.0%; Pred.No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 TGTGCTCTTCAGAAATCT 57
| | | | | | | | | | | | | | | | | |
DB 766 TGTGCTCTTCAGAAATCT 785

RESULT 23
US-10-027-632-256103
; Sequence 256103, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/219,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09

;
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 256103
; LENGTH: 1816
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-256103

Query Match 0.7%; Score 20; DB 13; Length 1816;
Best Local Similarity 100.0%; Pred.No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 717 AAATAAAAGGTTAAGAAA 736
| | | | | | | | | | | | | | | | | |
DB 1388 AAATAAAAGGTTAAGAAA 1407

RESULT 24
US-09-822-849A-13/c
; Sequence 13, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Pechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakr
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-03-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 2493
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1204-
; OTHER INFORMATION: n = a,c,t, or g
US-09-822-849A-13

Query Match 0.7%; Score 20; DB 9; Length 2493;
Best Local Similarity 100.0%; Pred.No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2476 ATTCCTGGCCACCTGGTCC 2495
| | | | | | | | | | | | | | | | | |
DB 564 ATTCCTGGCCACCTGGTCC 545

RESULT 25
US-09-901-136-3/c
; Sequence 3, Application US/09901136
; Publication No. US20030039968A1
; GENERAL INFORMATION:
; APPLICANT: HU, Song et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001273
; CURRENT APPLICATION NUMBER: US/09/901,136
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
```

2


```
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6505
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-6505

Query Match      0.7%; Score 20; DB 10; Length 577;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2505 AGAGGGTCCCAGAGATTAC 2524
    |||||
Db 345 AGAGGGTCCCAGAGATTAC 364

RESULT 16
US-09-263-959-372
; Sequence 372, Application US/09263959
; Patent No. US20020150891A1
; GENERAL INFORMATION:
; APPLICANT: Hood, Leroy E.
; APPLICANT: Rowen, Lee
; APPLICANT: Koop, Ben F.
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH UTI
; NUMBER OF SEQUENCES: 1279
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/263,959
; FILING DATE: 05-MAR-1999
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 920010.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 372:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-263-959-372

Query Match      0.7%; Score 20; DB 10; Length 701;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1842 CCAAAGGGAGAAATTGGC 1861
    |||||
Db 221 CCAAAGGGAGAAATTGGC 240

RESULT 17
US-09-393-982-61
; Sequence 61, Application US/09393634
```

```
; Patent No. US20020051997A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: SF, a No. US20020051997A1e1 Family of Taste Receptors
; FILE REFERENCE: 02307E-098000US
; CURRENT APPLICATION NUMBER: US/09/393,634
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human GR14
US-09-393-634e61

Query Match      0.7%; Score 20; DB 9; Length 954;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TGGTTTGGTGGCTGCTTCTT 48
    |||||
Db 386 TGGTTTGGTGGCTGCTTCTT 405

RESULT 18
US-09-510-332-27
; Sequence 27, Application US/09510332
; Publication No. US2003002278A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: T2R, a No. US2003002278A1e1 Family of Taste Receptors
; FILE REFERENCE: 02307E-098010US
; CURRENT APPLICATION NUMBER: US/09/510,332
; CURRENT FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 09/393,634
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: huamn T2R14 (hGR14)
US-09-510-332-27

Query Match      0.7%; Score 20; DB 11; Length 954;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TGGTTTGGTGGCTGCTTCTT 48
    |||||
Db 386 TGGTTTGGTGGCTGCTTCTT 405

RESULT 19
US-10-383-982-61
; Sequence 61, Application US/10383982
```

;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 10823
;; LENGTH: 479
;; TYPE: DNA
;; ORGANISM: Homo sapiens

;; FEATURE:
;; OTHER INFORMATION: MAP TO AC006518.17
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.94
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.99
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
US-09-864-761-10823

Query Match 0.7%; Score 20; DB 9; Length 479;
Best Local Similarity 100.0%; Pred.No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TGGTTTGGTGCTGCTCTT 48
|||||
Db 386 TGGTTTGGTGCTGCTCTT 367

RESULT 12
US-09-738-973-174/c

;; Sequence 174, Application US/09738973
;; Publication No. US20020110563A1
;; GENERAL INFORMATION:
;; APPLICANT: Reed, Steven G.
;; APPLICANT: Henderson, Robert A.
;; APPLICANT: Lodes, Michael J.
;; APPLICANT: Fling, Steven P.
;; APPLICANT: Mohamath, Raodoh
;; APPLICANT: Algate, Paul A.
;; APPLICANT: Secrist, Heather
;; APPLICANT: Indrias, Carol Yoseph
;; APPLICANT: Benson, Darin R.
;; APPLICANT: Elliot, Mark
;; APPLICANT: Mannion, Jane
;; APPLICANT: Kelos, Michael D.
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
;; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
;; FILE REFERENCE: 210121.475C9
;; CURRENT APPLICATION NUMBER: US/09/738,973
;; CURRENT FILING DATE: 2000-12-14
;; NUMBER OF SEQ ID NOS: 597
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 174
;; LENGTH: 548
;; TYPE: DNA
;; ORGANISM: Homo sapien
US-09-738-973-174

Query Match 0.7%; Score 20; DB 10; Length 548;
Best Local Similarity 100.0%; Pred.No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2476 ATTCCTGGGCCACCTGGTCC 2495
|||||
Db 543 ATTCCTGGGCCACCTGGTCC 524

RESULT 13

US-09-854-133-174/c
;; Sequence 174, Application US/09854133
;; Publication No. US20020183499A1
;; GENERAL INFORMATION:
;; APPLICANT: Lodes, Michael J.
;; APPLICANT: Mohamath, Raodoh
;; APPLICANT: Henderson, Robert A.
;; APPLICANT: Benson, Darin R.
;; APPLICANT: Secrist, Heather
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
;; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
;; FILE REFERENCE: 210121.475C10
;; CURRENT APPLICATION NUMBER: US/09/854,133
;; CURRENT FILING DATE: 2001-05-11
;; NUMBER OF SEQ ID NOS: 735
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 174
;; LENGTH: 548
;; TYPE: DNA
;; ORGANISM: Homo sapien
US-09-854-133-174

Query Match 0.7%; Score 20; DB 10; Length 548;
Best Local Similarity 100.0%; Pred.No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2476 ATTCCTGGGCCACCTGGTCC 2495
|||||
Db 543 ATTCCTGGGCCACCTGGTCC 524

RESULT 14

US-10-144-649A-174/c
;; Sequence 174, Application US/10144649A
;; Publication No. US20030118599A1
;; GENERAL INFORMATION:
;; APPLICANT: Lodes, Michael J.
;; APPLICANT: Wang, Tongtong
;; APPLICANT: Fan, Liqun
;; APPLICANT: Algate, Paul A.
;; APPLICANT: McNeill, Patricia D.
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
;; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
;; FILE REFERENCE: 210121.475C11
;; CURRENT APPLICATION NUMBER: US/10/144,649A
;; CURRENT FILING DATE: 2002-08-21
;; NUMBER OF SEQ ID NOS: 749
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 174
;; LENGTH: 548
;; TYPE: DNA
;; ORGANISM: Homo sapien
US-10-144-649A-174

Query Match 0.7%; Score 20; DB 14; Length 548;
Best Local Similarity 100.0%; Pred.No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2476 ATTCCTGGGCCACCTGGTCC 2495
|||||
Db 543 ATTCCTGGGCCACCTGGTCC 524

RESULT 15

US-09-867-70k-6505
;; Sequence 6505, Application US/09867701
;; Patent No. US20020132237A1
;; GENERAL INFORMATION:
;; APPLICANT: Aglate, Paul A.
;; APPLICANT: Jones, Robert
;; APPLICANT: Harlocker, Susan L.
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER

```
; Sequence 16328, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 16328
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034452.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.63
US-09-864-761-16328

Query Match 1.0%; Score 28; DB 9; Length 466;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2573 GACGTCCAGGTGTCAGAGGATTAAGG 2600
|||||
Db 466 GACGTCCAGGTGTCAGAGGATTAAGG 439

RESULT 10
US-09-938-842A-2425
; Sequence 2425, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Harber, Jeff
; APPLICANT: Kieps, Joel
; APPLICANT: Wang, Xue
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRP300-3
; CURRENT APPLICATION NUMBER: US/39/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,966
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2425
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2425

Query Match 0.7%; Score 20; DB 10; Length 393;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 746 TTTCACCAAAAGATATAA 765
|||||
Db 33 TTTCACCAAAAGATATAA 52

RESULT 11
US-09-864-761-10823/c
; Sequence 10823, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
```

```

; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31680
; LENGTH: 181
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034452.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 29
; OTHER INFORMATION: EST HUMAN HIT: N46880.1, EVALUE 1.00e-38
; OTHER INFORMATION: NT HIT: AF121217.1, EVALUE 1.70e+00
US-09-864-761-31680

Query Match          2.9%; Score 82; DB 9; Length 181;
Best Local Similarity 100.0%; Pred. No. 2.2e-31;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2598 AGGCTACCGAGAGAAATGGGAAAAGGAGCCAGGGTTTGGGTATCTCGAGAAC 2657
Db 165 AGGCTACCGAGAGAAATGGGAAAAGGAGCCAGGGTTTGGGTATCTCGAGAAC 106

Qy 2658 AGGTCTCTCGTGTCCCCAGGT 2679
Db 105 AGGTCTCTCGTGTCCCCAGGT 84

RESULT 7
US-10-027-632-124647/c
; Sequence 124647, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 197684
; LENGTH: 663
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-197684

Query Match          1.6%; Score 46; DB 13; Length 663;
Best Local Similarity 100.0%; Pred. No. 8.6e-13;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1804 GGAATCCAGGATTTCTTGGAAACCGAGGATTAATGGCCAAAGG 1849
Db 233 GGAATCCAGGATTTCTTGGAAACCGAGGATTAATGGCCAAAGG 188

RESULT 9
US-09-864-761-16323/c

```

```

; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 924647
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-124647

Query Match          2.0%; Score 57; DB 13; Length 1137;
Best Local Similarity 100.0%; Pred. No. 1.8e-18;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1955 AGGAAGCAAGGTGAACCTGGAATTCAGGGATGCTGGGCTTCTGGCTCAAGG 2011
Db 65 AGGAAGCAAGGTGAACCTGGAATTCAGGGATGCTGGGCTTCTGGCTCAAGG 9

RESULT 8
US-10-027-632-197684/c
; Sequence 197684, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 197684
; LENGTH: 663
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-197684

Query Match          1.6%; Score 46; DB 13; Length 663;
Best Local Similarity 100.0%; Pred. No. 8.6e-13;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1804 GGAATCCAGGATTTCTTGGAAACCGAGGATTAATGGCCAAAGG 1849
Db 233 GGAATCCAGGATTTCTTGGAAACCGAGGATTAATGGCCAAAGG 188

RESULT 9
US-09-864-761-16323/c

```

QY 2839 CCGTTCAGAAAGACCAAACTATTAG 2865
Db 325 CCGTTCAGAAAGACCAAACTATTAG 351

RESULT 4

US-09-918-995-33525
; Sequence 33525, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33525
; LENGTH: 412
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(412)
; OTHER INFORMATION: n = A,T,C or G

Query Match 7.8%; Score 224; DB 11; Length 412;
Best Local Similarity 100.0%; Pred. No. 1.2e-104;

Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1279 TGCCTTAATGTCCTCCAGTGATGTAGGTTCAACTCCAGTCCCTGATTTCTCCCGGA 1338
Db 188 TGCCTTAATGTCCTCCAGTGATGTAGGTTCAACTCCAGTCCCTGATTTCTCCCGGA 247
QY 1339 AAACGAGACTTCAAGGCCCAAGGTCACCTGAGTCTGCTGGGAACCTGGGTACCT 1393
Db 248 AAACGAGACTTCAAGGCCCAAGGTCACCTGAGTCTGCTGGGAACCTGGGTACCT 307
QY 1399 GGACAACTGTCAGATGGTAAAGCTGGATATCAGGAAATTCAGGACACACAGGTGT 1458
Db 308 GGACAACTGTCAGATGGTAAAGCTGGATATCAGGAAATTCAGGACACACAGGTGT 367
QY 1459 CCAGGATCTCCAGGAATACAGGAGCTCGAGGACTACAGGTTA 1502
Db 368 CCAGGATCTCCAGGAATACAGGAGCTCGAGGACTACAGGTTA 411

RESULT 5

US-09-864-761-32810/c
; Sequence 32810, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aomic-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49217
; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32810
; LENGTH: 221
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034452.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.63
; OTHER INFORMATION: NT HIT: L02917.1, EVALUE 2.10e+00
; OTHER INFORMATION: EST HUMAN HIT: AA451616.1, EVALUE 1.00e-103
; OTHER INFORMATION: SWISSPROT HIT: C53653, EVALUE 2.10e+03
US-09-864-761-32810

Query Match 5.8%; Score 166; DB 9; Length 221;
Best Local Similarity 100.0%; Pred. No. 1e-74;
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2435 GAAATTGTGATCATTTGCTGTCCTCCACATGCTCCCGGTATTCCTGGCCACCTGGTC 2494
Db 221 GAAATTGTGATCATTTGCTGTCCTCCACATGCTCCCGGTATTCCTGGCCACCTGGTC 162
QY 2495 CGATAGGCCCCAGAGGGTCCACAGGATTACCTGGTTTCCAGGAGAGATGGTGTTCCTG 2554
Db 161 CGATAGGCCCCAGAGGGTCCACAGGATTACCTGGTTTCCAGGAGAGATGGTGTTCCTG 102
QY 2555 GATTAGTGGGTGTCCTCGAGTCCAGTGTCCAGGATTAAGG 2600
Db 101 GATTAGTGGGTGTCCTCGAGTCCAGTGTCCAGGATTAAGG 56

RESULT 6

US-09-864-761-31680/c
; Sequence 31680, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aomic-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23

```

; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1e1 Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_rL_genes Version 1.0
; SEQ ID NO 282
; LENGTH: 2230
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (85)..(1152)
US-10-037-270-282

Query Match
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1062; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1804 GGAATCCAGGATTTCTCGAAACCGAGGATTAATGGCCAAAGGGAGAAATTTGGGCT 1863
Db 91 GGAATCCAGGATTTCTCGAAACCGAGGATTAATGGCCAAAGGGAGAAATTTGGGCT 150
QY 1864 CCAGGACAGCAAGGAAAAAAGAGCCCGGAGTGCCTGGTTTAAATGGGAAGCAATGGC 1923
Db 151 CCAGGACAGCAAGGAAAAAAGAGCCCGGAGTGCCTGGTTTAAATGGGAAGCAATGGC 210
QY 1924 TCACGAGGACGCTGGAAACACCGGGATCTAAGGAAGCAAGGTGAACCTGGAAATCAA 1983
Db 211 TCACGAGGACGCTGGAAACACCGGGATCTAAGGAAGCAAGGTGAACCTGGAAATCAA 270
QY 1984 GGGATCGCTGGGCTTCTGGGCTCAAGGAGAGAACGAGAGCAACGGGTTCGCCAGGAGAA 2043
Db 271 GGGATCGCTGGGCTTCTGGGCTCAAGGAGAGAACGAGAGCAACGGGTTCGCCAGGAGAA 330
QY 2044 CCAGGATACATGGTTTACCCTGGGATTCAGGAAAAAAGGGGACAAAGGAAATCAAGGT 2103
Db 331 CCAGGATACATGGTTTACCCTGGGATTCAGGAAAAAAGGGGACAAAGGAAATCAAGGT 390
QY 2104 GAAAAAGGTATTACGGGTCAAAAGGGAGAAAAATGGAAGACAGGGAAATTCAGGGCAACAG 2163
Db 391 GAAAAAGGTATTACGGGTCAAAAGGGAGAAAAATGGAAGACAGGGAAATTCAGGGCAACAG 450
QY 2164 GGAATTCAGGCCATCATGGTGCAAAAGAGAGAGAGGTGAAAAGGGAGAACCTTGGTGTG 2223
Db 451 GGAATTCAGGGCCATCATGGTGCAAAAGAGAGAGAGGTGAAAAGGGAGAACCTTGGTGTG 510
QY 2224 CGAGGTGCCATTGGATCAAAAGAGAGATCTGGGGTGGATGGCTTCATGGGGCCCGCAGGT 2283
Db 511 CGAGGTGCCATTGGATCAAAAGAGAGATCTGGGGTGGATGGCTTCATGGGGCCCGCAGGT 570
QY 2284 CTTAAGGGGCAACCTGGGGATCCAGTCTCTCAGGGACCCCGCAGGTTTGGATGGGAAGCCC 2343
Db 571 CTTAAGGGGCAACCTGGGGATCCAGTCTCTCAGGGACCCCGCAGGTTTGGATGGGAAGCCC 630
QY 2344 GGAAGAGGTTTTCAGAACAAATTTATTCAGCAAGTTTTCAGACAGATGTAATTAAGAGCCAG 2403
Db 631 GGAAGAGGTTTTCAGAACAAATTTATTCAGCAAGTTTTCAGACAGATGTAATTAAGAGCCAG 690
QY 2404 CTACCGAGTCTTACTTCAGAGTGGAGAAATTTAGAAATTTGATCATTTGCTCTCCCAACAT 2463
Db 691 CTACCGAGTCTTACTTCAGAGTGGAGAAATTTAGAAATTTGATCATTTGCTCTCCCAACAT 750
QY 2464 GGCTCCCGGGTATTCTCGGGCCACCTCGTCTCGCATAGGCCCGCAGAGGGTCCCGAGAGATTA 2523

```

90 18 0.6 557 13 US-10-027-632-85796 Sequence 85796, A
91 18 0.6 557 13 US-10-027-632-320345 Sequence 320345,
92 18 0.6 557 13 US-10-027-632-320346 Sequence 320346,
93 18 0.6 557 13 US-10-027-632-320347 Sequence 320347,
94 18 0.6 557 13 US-10-027-632-320348 Sequence 320348,
95 18 0.6 570 13 US-10-027-632-115045 Sequence 115045,
96 18 0.6 570 13 US-10-027-632-236526 Sequence 236526,
97 18 0.6 581 13 US-10-027-632-70650 Sequence 70650, A
98 18 0.6 583 13 US-10-027-632-73394 Sequence 73394, A
99 18 0.6 583 13 US-10-027-632-189884 Sequence 189884,
100 18 0.6 583 13 US-10-027-632-189885 Sequence 189885,

ALIGNMENTS

RESULT 1
US-09-764-864-232
; Sequence 232, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PIZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 232
; LENGTH: 2375
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-232

Query Match 44.7%; Score 1282; DB 10; Length 2375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1584 AAAGGTGAAATGGTGCCAAAGAGACAAAGGATCACCTGGATTATGGCAAAAGGG 1643
DB 9 AAAGGTGAAATGGTGCCAAAGAGACAAAGGATCACCTGGATTATGGCAAAAGGG 58
QY 1644 TGCAAAAGGTGAAAGGGGAATGCTGGCTTCCCTGGCCCTCCCTGGACCTGCTGGAGAAC 1703
DB 69 TGCAAAAGGTGAAAGGGGAATGCTGGCTTCCCTGGCCCTCCCTGGACCTGCTGGAGAAC 128
QY 1704 AGGAGACATGGAAGGATGATTAATGGGTAGTCCGGTTTCAAGGGAGACGAGATC 1763
DB 129 AGGAGACATGGAAGGATGATTAATGGGTAGTCCGGTTTCAAGGGAGACGAGATC 138
QY 1764 CCCTGGTGCTCCGGGCGAGGATGGAACACCGGGAGAGCCTGGAAATCCAGGATTTCTGG 1823
DB 189 CCCTGGTGCTCCGGGCGAGGATGGAACACCGGGAGAGCCTGGAAATCCAGGATTTCTGG 248
QY 1824 AAACCGAGGATTAATGGGCCAAAGGAGAGAAATGGGCTCCAGGACAGCAAGGAAAAA 1883
DB 249 AAACCGAGGATTAATGGGCCAAAGGAGAGAAATGGGCTCCAGGACAGCAAGGAAAAA 308
QY 1884 AGGAGCCCCAGGATGCTGTTTAAATGGGAAGCAATGGCTCACCGGCCAGCCTCGAAC 1943
DB 309 AGGAGCCCCAGGATGCTGTTTAAATGGGAAGCAATGGCTCACCGGCCAGCCTCGAAC 368
QY 1944 ACCGGGATCTAAGGGAAGCAAGGTGAACCTGGAAATCAAGGGATGCTGGGGCTTCTGG 2203
DB 369 ACCGGGATCTAAGGGAAGCAAGGTGAACCTGGAAATCAAGGGATGCTGGGGCTTCTGG 428
QY 2004 GCTCAAGGGAGAACCGAGCAACCGGTTCCCGAGGACACGAGTACATGGGTTTACC 2063
DB 429 GCTCAAGGGAGAACCGAGCAACCGGTTCCCGAGGACACGAGTACATGGGTTTACC 488
QY 2064 CGGGATTCAAGGAAAAAGGGGACAAAGGAATCAAGGTGAAAGGATTCAGGGTCA 2123

DB 489 CCGGATTCAAGGAAAAAGGGGACAAAGGAATCAAGGTCAAAAAGGTATTCAAGGTCA 548
QY 2124 AAAGGGAGAAATCGPAGACAGAGGAATTCAGGCGCACACGGGAATTCAGGCCATCATGG 2183
DB 549 AAAGGGAGAAATCGAAGACAGAGGAATTCAGGCGCAACAGGGAAATTCAGGCCATCATGG 608
QY 2184⁶ TGCAAAAGGAGAGAGAGGTGAAAGGAGAAACCTGGTGTCCGAGGTGCCATTGGATCAAA 2243
DB 609 TGCAAAAGGAGAGAGAGGTGAAAGGAGAAACCTGGTGTCCGAGGTGCCATTGGATCAAA 668
QY 2244 AGGAGAAATCTGGGGTGGATGCTTGTATGGGGCCCGCAGGTCTTAAGGGGCAACTGGGGA 2303
DB 669 AGGAGAAATCTGGGGTGGATGCTTGTATGGGGCCCGCAGGTCTTAAGGGGCAACTGGGGA 728
QY 2304 TCCAGGTCTCTCAGGACCCCGCAGGTTTGGATGGGAACCCGGAAGAGAGTTTTCAGAAACA 2363
DB 729 TCCAGGTCTCTCAGGACCCCGCAGGTTTGGATGGGAACCCGGAAGAGAGTTTTCAGAAACA 788
QY 2364 ATTTATTCGACAAAGTTTGCACAGATGTAATAAGAGCCAGCTACCACTCTTACTTCAGAG 2423
DB 789 ATTTATTCGACAAAGTTTGCACAGATGTAATAAGAGCCAGCTACCACTCTTACTTCAGAG 848
QY 2424 TGGAGAAATAGAAATTTGTGATCATTCGCTGTCCCAACATGGCTCCCGGGGTATTCCTGG 2483
DB 849 TGGAGAAATAGAAATTTGTGATCATTCGCTGTCCCAACATGGCTCCCGGGGTATTCCTGG 908
QY 2484 GCCACCTGGTCCGATAGCCCGCAGAGGGTCCCAAGAGGATTACCTGGTTTGCAGGAAAGAGA 2543
DB 909 GCCACCTGGTCCGATAGCCCGCAGAGGGTCCCAAGAGGATTACCTGGTTTGCAGGAAAGAGA 968
QY 2544 TGGTGTCTCTGGATTAGTGGTGTCCCTGGACGTCAGGTGTCAGAGGATTAAGAAGCCT 2603
DB 969 TGGTGTCTCTGGATTAGTGGTGTCCCTGGACGTCAGGTGTCAGAGGATTAAGAAGCCT 1028
QY 2604 ACCAGGAAGAAATCGGAAAGGAGGCAAGGGTTTGGGTATCTCCGAGAAACAAGTCC 2663
DB 1029 ACCAGGAAGAAATCGGAAAGGAGGCAAGGGTTTGGGTATCTCCGAGAAACAAGTCC 1088
QY 2664 TCCTGGTCTCCCGAGGTCAGAGGCCCCCTCTGGAAATAGCAAGAGAGTCTCTCCAGGAGA 2723
DB 1089 TCCTGGTCTCCCGAGGTCAGAGGCCCCCTCTGGAAATAGCAAGAGAGTCTCTCCAGGAGA 1148
QY 2724 CCGAGGTCTCTCGCAAGATGAGACCATGGAAACCTGGAAATCCAGGCGCAACAGG 2783
DB 1149 CCGAGGTCTCTCGCAAGATGAGACCATGGAAACCTGGAAATCCAGGCGCAACAGG 1208
QY 2784 CCCCCAGGATCTCGACCCCATCATATGTTTGTAGTGAATTCGCCAGAGAGATCCGTT 2843
DB 1209 CCCCCAGGATCTCGACCCCATCATATGTTTGTAGTGAATTCGCCAGAGAGATCCGTT 1268
QY 2844 CAGAAAGGACCAAACTATTAG 2865
DB 1269 CAGAAAGGACCAAACTATTAG 1290

RESULT 2

US-10-037-270-282
; Sequence 282, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyang
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aildong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 23:50:00 ; Search time 2728 Seconds
(without alignments)

2678.968 Million call updates/sec

Title: US-09-996-611B-5

Perfect score: 2865

Sequence: 1 aggtctactattacatt.....gaaaggagccaactattag 2865

Scoring table: OLIGO_NUC
Gapop 60.0 , Capext 60.0

Searched: 1708419 seqs, 1275411651 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3416838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Published Applications NA.*
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/prodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/prodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMAR:ES

Result No.	Score	Query Match	Length	DB ID	Description
1	1282	44.7	2375	10	US-09-764-864-232
2	1062	37.1	2230	14	US-10-037-270-282
3	267	9.3	496	11	US-09-918-995-17490
4	224	7.8	412	11	US-09-918-995-13525
5	166	5.8	221	9	US-09-864-761-32810
6	162	2.9	181	9	US-09-864-761-31680
7	57	2.0	1197	13	US-10-027-632-124647
8	46	1.6	663	13	US-10-027-632-197684
9	28	1.0	466	9	US-09-864-761-16328
10	20	0.7	393	10	US-09-938-842A-2425
11	20	0.7	479	9	US-09-864-761-10823
12	20	0.7	548	10	US-09-738-973-174
13	20	0.7	548	10	US-09-854-133-174
14	20	0.7	548	14	US-10-144-649A-174
15	20	0.7	577	10	US-09-867-701-6505
16	20	0.7	701	10	US-09-263-959-372

17	20	0.7	954	9	US-09-393-634-61	Sequence 61, Appl
18	20	0.7	954	11	US-09-510-332-27	Sequence 27, Appl
19	20	0.7	954	12	US-10-383-982-61	Sequence 61, Appl
20	23	0.7	1287	9	US-09-815-242-3957	Sequence 3957, Ap
21	20	0.7	1299	9	US-09-815-242-6757	Sequence 6757, Ap
22	20	0.7	1635	10	US-09-070-927A-621	Sequence 62, App
23	20	0.7	1816	13	US-10-027-632-256103	Sequence 256103,
C 24	20	0.7	2493	9	US-09-922-849A-13	Sequence 13, Appl
C 25	20	0.7	378361	11	US-09-901-136-3	Sequence 3, Appl
C 26	20	0.7	684973	10	US-09-263-959-1	Sequence 1, Appl
C 27	19	0.7	391	10	US-09-783-530-8616	Sequence 8616, Ap
C 28	19	0.7	464	11	US-09-928-995-30341	Sequence 30341, A
C 29	19	0.7	580	13	US-10-027-632-28811	Sequence 28811, A
C 30	19	0.7	998	12	US-09-814-353-20454	Sequence 20454, A
C 31	19	0.7	1146	14	US-10-096-162-1	Sequence 1, Appl
C 32	19	0.7	1188	13	US-10-027-632-117674	Sequence 117674,
C 33	19	0.7	1188	13	US-10-027-632-117675	Sequence 117675,
C 34	19	0.7	1188	13	US-10-027-632-117676	Sequence 117676,
C 35	19	0.7	1188	13	US-10-027-632-123666	Sequence 23666,
C 36	19	0.7	1300	14	US-10-184-644-269	Sequence 269, App
C 37	19	0.7	1300	14	US-10-184-644-269	Sequence 269, App
C 38	19	0.7	1363	10	US-09-938-842A-3872	Sequence 3872, Ap
C 39	19	0.7	2341	12	US-10-240-965-152	Sequence 152, App
C 40	19	0.7	2346	14	US-10-125-237-52	Sequence 52, Appl
C 41	19	0.7	2346	14	US-10-105-891-52	Sequence 52, Appl
C 42	19	0.7	2652	13	US-10-027-632-111629	Sequence 111629,
C 43	19	0.7	2652	13	US-10-027-632-111630	Sequence 111630,
C 44	19	0.7	3609	14	US-10-128-714-386	Sequence 386, App
C 45	19	0.7	3609	14	US-10-128-714-386	Sequence 386, App
C 46	19	0.7	66686	10	US-09-736-960-86	Sequence 86, Appl
C 47	18	0.6	25	24	US-10-098-263B-48341	Sequence 48341, A
C 48	18	0.6	115	9	US-09-815-242-3122	Sequence 3122, Ap
C 49	18	0.6	118	9	US-09-864-761-21267C	Sequence 21267C, A
C 50	18	0.6	139	9	US-09-815-242-3611	Sequence 3611, Ap
C 51	18	0.6	186	10	US-09-783-590-12193	Sequence 12193, A
C 52	18	0.6	261	9	US-09-815-242-3045	Sequence 3045, Ap
C 53	18	0.6	264	13	US-10-040-739-74C	Sequence 740, App
C 54	18	0.6	304	13	US-10-040-739-468	Sequence 468, App
C 55	18	0.6	340	10	US-09-954-456-1386	Sequence 1386, Ap
C 56	18	0.6	340	10	US-09-954-531-1136	Sequence 1136, Ap
C 57	18	0.6	372	9	US-09-815-242-3153	Sequence 3153, Ap
C 58	18	0.6	377	10	US-09-736-457-1127	Sequence 1127, Ap
C 59	18	0.6	377	10	US-09-736-457-1136	Sequence 1136, Ap
C 60	18	0.6	377	10	US-09-902-941-1127	Sequence 1127, Ap
C 61	18	0.6	377	10	US-09-902-941-1136	Sequence 1136, Ap
C 62	18	0.6	377	10	US-09-849-626-1127	Sequence 1127, Ap
C 63	18	0.6	377	10	US-09-849-626-1136	Sequence 1136, Ap
C 64	18	0.6	377	12	US-10-113-872-1127	Sequence 1127, Ap
C 65	18	0.6	377	12	US-10-113-872-1136	Sequence 1136, Ap
C 66	18	0.6	377	14	US-10-017-754-1127	Sequence 1127, Ap
C 67	18	0.6	377	14	US-10-017-754-1136	Sequence 1136, Ap
C 68	18	0.6	394	13	US-10-027-632-52921	Sequence 52921, A
C 69	18	0.6	395	13	US-10-027-632-82160	Sequence 82160, A
C 70	18	0.6	395	13	US-10-027-632-302139	Sequence 302139,
C 71	18	0.6	429	11	US-09-918-995-5943	Sequence 5943, Ap
C 72	18	0.6	439	11	US-09-918-995-13940	Sequence 13940, A
C 73	18	0.6	441	10	US-09-954-456-327	Sequence 327, App
C 74	18	0.6	441	10	US-09-954-456-327	Sequence 327, App
C 75	18	0.6	457	11	US-09-918-995-21550	Sequence 21550, A
C 76	18	0.6	461	9	US-09-864-761-5910	Sequence 5910, Ap
C 77	18	0.6	467	9	US-09-770-444-259	Sequence 259, App
C 78	18	0.6	469	9	US-09-815-343-359	Sequence 359, App
C 79	18	0.6	474	9	US-09-864-761-13995	Sequence 13995, A
C 80	18	0.6	479	11	US-09-918-995-23654	Sequence 23654, A
C 81	18	0.6	515	9	US-09-864-761-7318	Sequence 7318, A
C 82	18	0.6	517	13	US-10-027-632-87326	Sequence 87326, A
C 83	18	0.6	517	13	US-10-027-632-367167	Sequence 367167,
C 84	18	0.6	531	11	US-09-862-540-18	Sequence 18, Appl
C 85	18	0.6	550	13	US-10-027-632-193269	Sequence 193269,
C 86	18	0.6	550	13	US-10-027-632-193270	Sequence 193270,
C 87	18	0.6	557	13	US-10-027-632-85793	Sequence 85793, A
C 88	18	0.6	557	13	US-10-027-632-85794	Sequence 85794, A
C 89	18	0.6	557	13	US-10-027-632-85795	Sequence 85795, A

Db	4346	CGGTCCTCAGGGACCTGCAGGAAAGCCTGGTCCAGA-----AGGTCTTCGGGGC	4394
Qy	2447	ATTGCCCTGTCCCAACATGGCTCCCGGGTATTCTCTGGGCCACCTGGTCCCGATAGGCCCCAG	2506
Db	4395	ATCCCTGGTCC--TGTGGGAGAACAAAGTCTCCCTGGAGCTGCAGGCCAAGATGGACCAC	4452
Qy	2507	AGGTCCCCAGAGGATTACCTGGTTTTCAGGAGAGAGATGTTCTCTGGATTTAGTGGGTG	2566
Db	4453	CTGGTCTCTATGGGACCTCCTGGCTTACCTGGTCTCAAAAGGTGACCTGGCTCCAAAGGGTG	4512
Qy	2567	TCCCTGGACGTCCAGGTGTCAAGGATTAAANGCCTACCAGGAGAAATGGGGAAAAAG	2626
Db	4513	AAAAAGGACATCCTGGTTTAAATTGGCCTGATTGGTCTCTCCAGGAGAAACAAGGGGAAAAAG	4572
Qy	2627	GGAGCCAGGGTTTGGGTATCTCTGGAGAACAAAGGTCTCTCTGGTCCCGCCAGTCCAGAGG	2686
Db	4573	GTGACCGAGGGCTC-----CCTGGAATCAAGGATCTCCAGGAGCAAAAGGGGATGGGG	4626
Qy	2687	GCCCTCTCTGGAATTAAGCAAGAGGTCTCTCAGGAGACCCAGGTCTCCCTGGGCAAGATG	2746
Db	4627	GAATTCTCTGGTCTT-----GCTGGTCCCTTAGGTCCACCTGGTCTCTCCAGGCTTACCAG	4680
Qy	2747	GAGACCATGGAAAAACCTGGGAATCCAAAGGCCAACAGGCCCGCCCGAGGC	2793
Db	4681	GTCTCAAGGCCCAAGGGTAAACAAGGCTCTACTGGACCGCTGGC	4727

Search completed: October 3, 2003, 00:00:11
Job time : 715 secs

Db 4346 CGGTCTCAGGACCTGACGAAAGCCTGTGTCAGA-----AGTCTTCGGGGC 4394
Qy 2447 ATTGCTGTCCCAACATGGCTCCCGGGGTATTCTCGGCCACCTGGTCCGATAGGCCAG 2506
Db 4395 ATCCCTGGTCC--TGTGGGAAACAGGTCTCTCTGGAGCTGACGGCCAGATGGACAC 4452
Qy 2507 AGGTCCAGAGGATTACTGGTTTGCAGAGAGATGGTGTCTCTGGATTAGTGGGTG 2566
Db 4453 CTGTCTCTATGGGACCTCTGGCTTACCTGTCTCAAAAGGTGACCTGGCTCCAAAGGTG 4512
Qy 2567 TCCCTGGACGCTCAGGCTCAGAGGATTAAAGGCTACCGAGGAGAAATGGGGAAGAAG 2626
Db 4513 AAAAGGACATCTCTGGTTTAAATTGGCTGATTGGTCTCTCAGGAGAAACAAAGGGGAAAAG 4572
Qy 2627 GGAGCCAAAGGTTTGGGTATCTCTGGAGAACAAAGTCTCTCTGGTCCCGCCAGGTCCAGAGG 2686
Db 4573 GTGACCGAGGCTC-----CCTGGAACTCAAGGATCTCCAGGAGCAAAAGGGGATGGG 4626
Qy 2687 GCTCTCTGGAAATAGCAAGAGTCTCTCAGAGACCCAGGTCTCTCTGGGCAAGATG 2746
Db 4627 GAATTCCTGTCTCT-----GCTGGTCTCTTAGGTCTCACTGTCTCTCCAGGCTTACAG 4680
Qy 2747 GAGACCATGAAAACCTCGAATCAAGGGCAACAGGCCCGCCCGCAGGC 2793
Db 4691 GCTCAAGGCCCAAGGTTAAACAAAGGCTCTACTGGACCCGCTGGC 4727

RESULT 15

US-09-954-456-762
; Sequence 762, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 762
; LENGTH: 6158
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: n=a,t,g or c
US-09-954-456-762

Query Match 8.4%; Score 239.8; DB 1C; Length 6158;

Best Local Similarity 50.8%; Pred. No. 5.4e-55;
Matches 756; Conservative 0; Mismatches 687; Indels 44; Gaps 6;

Qy 1326 TTGTCTCCGGGAAACACGAGACTTCAAGGGCCCAAGGTGACCTGACTGCTGCTGGAA 1385
Db 3266 TGGAGCTCAGGGTGCACCTGGACTGAAAGGAGGGGAAGTCCCCAGGGCCACACAGGTCC 3325
Qy 1386 CCTTGGCTACCTTGACAACTGGTCAAGATGGTAAAGCTTGATATCAGGGAATTCAGG 1445
Db 3326 AGTTGGCTCACACAGGAAACGTGGGTACAGAGGTACAGCTGGCCCAATTTGGTTACGAGG 3385
Qy 1446 GACACACAGTGTTCAGGATCTCCAGGAATACAGAGAGCTCGAGACTACACAGGTATCAA 1505
Db 3386 GGGCCCGGACCTCAGSGTCTCTCTGGTCCAGCTGAGAGAAAGTGTCTCTCGAGAAAA 3445
Qy 1506 AGGAAACACAGGGCAGATGTGACAAAGGTGATGCTGGACTTCTTGGTTTCTCTGGGCT 1565
Db 3446 AGTCCCCAAAGGCTCTCAGGAGAGATGGAGTCAAGTCTCTGTGTCTCTCCAGGGCC 3505
Qy 1566 CATGGCATCCAGGATCAAGAGGTGAAGTGGTCCAAAGGAGACAAAGATCACCTGG 1625
Db 3506 AGCTGGTCTCTCGGCTCCCTGGGAAAGACGAGACAAAGGTGAAATGGTAGAGCCGG 3565
Qy 1626 ATTTATGGCAAAAGGCTGCAAAAGGTGAAAGGGGAATGCTGCTTCCCTGGCTCTCC 1685
Db 3566 ACAAAAGGACAGAGGTGGCAAGGAGAAATGGCCCTCCCGTCCCCAGGCTTCA 3625
Qy 1686 TGGACTCTCTGGAGAACACAGGAGACATGGAAGATGGAATTAATGGGTAGTCCCGGTTT 1745
Db 3626 AGGACCAAGTTGTGCTCCCTGGAATTTCTGGAGGTGATGTTGAACACAGTCTCTAGAGACA 3685
Qy 1746 CAAGGAGAACAGCATCCCTGGTGTCTCCGGGACAGATGGACACAGGGAGAGGCTGG 1805
Db 3686 GCAGGGGATGTTTGGGCAAAAGGTGATGAGGTGCCAGAGGCTTCCCTGGACCTCTCG 3745
Qy 1806 AATCCAGAGATTCTCGAAACCCAGGATTAATGGGCCAAAGGGAGAAAT----- 1857
Db 3746 TCCAATAGTCTTCAAGGTCTGTCAGGCCCCACCTGTTGAAAAAGGTGAAATGGGATGT 3805
Qy 1858 -----GGGCTCTCAGACACAGCAAGGAAAAAGGAGCCCGGAGTGCCTGGTT 1907
Db 3806 TGGTCCAATGGGGGCCACCTGGTCTCCAGGGCCCAAGAGGCCCTCAAGTCCCAATGGAGC 3865
Qy 1908 AATGGAGCAATGCTCACACAGGCCACCTGGGAACACCGGATCTAAGGAAACAAAGG 1967
Db 3866 TGATGGACCAACAGGACCCCGAGTCTCTGTTGGTTCAGTTGGTGTGGTGTGGAGAAAGG 3925
Qy 1968 TGAACCTGGAATCAAGGATGCTCTGGGCTTCTGGGCTCAAGGAGAACACAGAGCAAC 2027
Db 3926 TGAACCTGGAGAACGAGGAAACCCAGGGCCCTCTGGGAAAGCAGGTGTAGCGGTCCCA 3985
Qy 2028 GGGTTCCCGCAGGAGAACAGGATACATGGGTTTACCCGGGATTCAGGAAAAAGGGGA 2087
Db 3986 AGGAAAGAGAGAGAGAAAGGGAAAGCTGGTCCACCTGGAGCTGCTGGACCTCCAGGTGC 4045
Qy 2089 CAAGGAAATCAAGGTGAAAAAGGTATTTCAGGGTCAAAAGGAGGAAAAATGGAGACAGGG 2147
Db 4046 CAAGGGCGCCAGGTGATGATGGCCCTAAGGTTAAACCCGGTCTCTGTGTTCTCTGG 4105
Qy 2148 AATTCAGGGCAACAGGAAATCAAGGCAATCATGGTCAAAAGGAGAGAGAGGTGAAA 2207
Db 4106 AGATCTGTCTCTCTGGGAACTTGGGCCCTGCAAGGTCAAGATGGTGTGGTGTGACA 4165
Qy 2208 GGGAGAACTGTGTGCCAGGTGCCATTTGGATCAAAAGGAGAAATCTGGGTGGATGGCTT 2267
Db 4166 GGTGAAATGGAGATCTGTCAACCGGTCTCTCTGGGCCCATCTGGTAGGCTTGCCCC 4225
Qy 2268 GATGGGCCCGCAGTCTCTAAGGGGCAACCTGGGATCCAGGTCTCTCAGGACCCCCAGG 2327
Db 4226 ACCAGTCTCTCTGAAAAACGAGTCTCTCTCTGGAGCTGCAGGTGCAGAGGAGAGCAAGG 4285
Qy 2328 TTGTGATGGGAAGCCCGAAGAGAGTTTTCAGAACAAATTTATTCGACAAAGTTTTCACAG- 2386
Db 4286 TGAAAAAGGTCTTAAGGGGGAAGCAGGTGACAGAGGTCTCTCTGGAAAAACCGGCCAGT 4345
Qy 2387 ATGTAAATAGAGCCCGAGCTACCACTCTTACTTCAGAGTGGAGAAATAGAAATTTGTATC 2446

Db 4263 AGATCCTGGTCTCTCTGGGAACTTGGCCCTGCAGCTCAAGATGGTGTGGTGTGACAA 4322
 QY 2208 GGGAGAACTGGTCTCCAGAGTGCATTTGGATCAAAAGGAAATCTGGGGTGGATGCTT 2267
 Db 4323 GGGTGAAGATGAGATCTGGTCAACCGGGTCTCTGGCCCATCTGGTGAAGTGGCCC 4382
 QY 2268 GATGGGGCCGAGATCTTAAGGGGCAACCTTGGGATCCAGGTCTCAGGGACCCGAGG 2327
 Db 4383 ACCAGTCTCTCTGAAACAGAGTCTCTGGAGTCCAGTGCAGAGGGAACAGG 4442
 QY 2328 TTGGATGGAGAGCCCGGAGAGAGATTTTCAGACAAATTTATTCGACAAAGTTTCACAG- 2386
 Db 4443 TGAATAAGGTGCTAAGGGGGAAGCAGAGTGCAGAGGTCTCTCGAATAACCGGCCAGT 4502
 QY 2387 ATGTAATAAGAGCCAGCTACAGGTCTTACTTCAGAGTGAAGAAATAGAAATTTGTATC 2446
 Db 4503 CGTCTCAGGACCTCAGGAAAGCCTGGTCCAGA-----AGGTCTCGGGGC 4551
 QY 2447 ATTGCTGTCCCAACATGGCTCCCGGGTATTCTTGGGCCACCTGGTCCGATAGGCCAG 2506
 Db 4552 ATCCCTGGTCC--TGTGGAGAAACAAGTCTCCCTGGAGCTGCAGGCCAAGATGACCCAC 4609
 QY 2507 AGGTCCAGAGGATTAAGTCTTGGTCCAGGAGAGATGGTCTCTGGATTTAGTGGGTG 2566
 Db 4610 CTGGTCTATGGAGCTCTGGCTTACCTGGTCTCAAGGTGACCCCTGGCTCCAGAGGTG 4669
 QY 2567 TCCCTGGAGCTCCAGGTGTACAGATTAAGAGGCTTACAGGAGAAATGGGAAAAAG 2626
 Db 4670 AAAGGGACATCTGTTTAAATGGCTGATTGGTCTCCAGGAGAACAGGGGAAAG 4729
 QY 2627 GGAGCCAGGTTTGGGTATCTCTGAGAACAGGTCTCTCTGGTCCCGCCAGGTCCAGAGG 2686
 Db 4730 GTGACCGAGGCTC-----CCTGGAATCAAGGATCTCCAGGAGCAAAAGGGATGGG 4783
 QY 2687 GCCCTCTGGAATAGCAAGAGTCTCTCCAGGAGACCCAGGTCTCCTCGCAAGATG 2746
 Db 4784 GAATCTCTGGTCT-----GCTGTCTCTTAGTCTACCTGGTCTCTCCAGGCTTACAG 4837
 QY 2747 GAGACCATGGAACCTCGAATCCAAAGGCAACCAAGGCCCGCCCGCCAGGC 2793
 Db 4838 GTCTCAAGGCCCAAGGGTAACAAAGGCTCTACTGGACCCGCTGGC 4884

RESULT 14
 US-09-919-497-6
 ; Sequence 6, Application US/09/19497
 ; Patent No. US20020106662A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Muller, George L.
 ; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
 ; FILE REFERENCE: B0801/7225
 ; CURRENT APPLICATION NUMBER: US/09/919,497
 ; CURRENT FILING DATE: 2001-07-31
 ; PRIOR APPLICATION NUMBER: US 60/221,735
 ; PRIOR FILING DATE: 2000-07-31
 ; NUMBER OF SEQ ID NOS: 100
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 6158
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: Unsure
 ; LOCATION: (2434)..(2434)
 ; OTHER INFORMATION: n = a, c, g or t/u
 US-09-919-497-6
 Query Match 8.4%; Score 239.8; DB 10; Length 6158;
 Best Local Similarity 50.8%; Pred. No. 5.4e-55;
 Matches 756; Conservative 0; Mismatches 687; Indels 44; Gaps 6;
 QY 1326 TTGTCTCCGGGAAACCAAGGACTTCAAGGCCCAAGGTGACCCCTGGACTGCCTGGAA 1395

0
 Db 3266 TGAGCTCAGGCTGCACTGACTGAAAGAGGGGAAAGTCCCAAGGCCCAACAGGTCC 3325
 QY 1386 CCTGGCTACCTGGNCAACCTGGTCAAGATGGTAAGCTGGATATCAGGGAATTGCAGG 1445
 Db 3326 AGTTGGCTCACCAGGAGAACGTTGGGTACAGAGGTACAGCTGCCCAATTTGGTTACGAGG 3385
 QY 1446 GACACCAGGTGTTCAGGATCTCCAGGAATACAAGAGCTCCAGGACTACCAAGGTATACAA 1505
 Db 3386 GCGCCCGGACCTCAGGTCTCTCTGGTCCAGCTGGAGAGAAAGTGTCTCTTGGAGAAAA 3445
 QY 1506 AGAGAACCCAGGCGAGATGGTGAACAAGGTGATCTGTGACTTCTGTGTTTCTTGGGT 1565
 Db 3446 AGTCTCCCAAGGCTCTGCAGGAGAGATGAGTTCAAGGTCTCTGTCTCCAGGGGCC 3505
 QY 1566 TCATCGCATGCCAGGATCAAGGGTGAATGGTCCCAAGGAGCAAGAGATCACTGG 1625
 Db 3506 AGTGGTCTCTCGGCTCCCTGGGGAAGACCGAGACAAGGTGAATTTGGTGAAGCCGG 3565
 QY 1626 ATTTATGGCAAAAAGGTGCAAAAAGTGAAGGGGAATGCTGGCTTCCCTGGGCTCC 1685
 Db 3566 ACAAAGAGCAGCAAGGGTGGCAAGGAGAAAAATGGCCCTCCCGTCCCAAGGTCTTCA 3625
 QY 1686 TSGACCTGCTGAGAACCCAGGAGACATGGAAGATGATTAATGGGTAGTCCCGTTT 1745
 Db 3626 AGGACAGTTGGTGGCCCTGGAAATTTGCTGAGGTGATGGTGAACCCAGGTCTTAGAGACA 3685
 QY 1746 CAAAGGAGAACGAGGATCCCTGGTCTCCGGGAGGATGAAACACCGGGAGAGCCTGG 1805
 Db 3686 GCAGGGATGTTTGGGCAAAAAGGTGATGAGGTGCCAGAGGCTTCCCTGGACCTCTGG 3745
 QY 1806 AATCCAGGATTTCTCTGGAAACCGAGGATTAATGGGCCAAAAGGGAGAAAT----- 1857
 Db 3746 TCCATATAGTCTTCAAGGTCTCCAGGCCCACTGGTGAAGGGTGAAGGTGGGGATGT 3805
 QY 1859 -----GGCCCTCCAGGACAGCAAGGAAAAAGAGCCCGGATCTAAGGGAGCAAGG 1967
 Db 3806 TGGTCTCTGGGGCCCACTGGTCTCCAGGCCCAAGAGGCCCTCAAGGTCCCAATTTGGAGC 3865
 QY 1968 AATGGGAAGCAATGGCTCACAGGCGAGCTGGAACACCGGATCTAAGGGAGCAAGAGG 1967
 Db 3866 TGAATGACCAAGGACCCCGAGGTCTGTGTGTTTCAAGTGTGTGTGGAGAAAGG 3925
 QY 1968 TGAACCTGGAATTCAGGATGCTGGGGCTCTGGGCTCAAGGAGAACCCAGGAGCAAC 2027
 Db 3926 TGAACCTGGAAGCAGGAAACCCAGGCTCTCTGGGAAAGCAGGTAGGGGCTCCCAA 3985
 QY 2028 GGGTCTCCCGAGGAGAACAGGATACATGGGTTTACCCGGATTCAGGAAAAAGGGGA 2087
 Db 3986 AGGAGAAAGAGGAGAAAGGGAGCTGTCCACTTGGAGCTGTGGACCTCCAGGTGC 4045
 QY 2088 CAAAGGAAATCAAGGTGAAAAAGTATTCAAGGTCAAAAGGGAGAAAAATGGAAGACAGG 2147
 Db 4046 CAAAGGGCCCGCAGGTGATGATGGCCCTAAGGGTAACCCGGGTCTGTGTGTTTCTCTGG 4105
 QY 2148 AATTCAGGGCAACAGGAAATCAAGGCAATCATGTGCAAAAGGAGAGAGAGGTGAAGA 2207
 Db 4106 AGATCTCTCTCTCTGGGAACTTGGGCCCTCAGGTCAAGATGTGTGTGTGTGTGACAA 4165
 QY 2208 GGGAGAACCTGTGTCCGAGGTGCCATTGGAATCAAAAGGAGAACTCTGGGTGGATGGT 2267
 Db 4166 GGTGAAATGAGATCTGTGTCAACCGGCTCTCTGGCCCTCTGTGAGCTGGGCC 4225
 QY 2268 GATGGGGCCCGAGGTCTCTAAGGGCAACCTGGGGATCCAGGTCTCTAGGGACCCCGAGG 2327
 Db 4226 ACCAGGTCTCTCTGGAAAAACGAGGCTCTCTGAGCTGCAGGTGCAGAGGGAAGCAAGG 4285
 QY 2328 TTTGGATGGAAAGCCCGGAGAGAGTTTTCAGAACANTTTATTCAGCAAGTTTGCACAG- 2386
 Db 4286 TGAATAAGGTGCTAAGGGGGGAAGCAGGTGCAAGAGGTCTCTCTGGAAAAAACCGGCCAGT 4345
 QY 2387 ATCTAATAGACCCAGCTACAGGTCTTACTTCAGGTGGAAGAAATTAGMAATTGTGATC 2446

```

QY 1928 CAGCCAGCCTGGAAACACCGGATCTAAGGAGCAAGTGAACCTGGAAATTCAGGGG 1987
Db 1271 CTGCGCCCTCTGGGATTAATGGTGTCTGTGTAAAGCGAAATGGGTCCCGCTGGCA 1330
QY 1988 TGCCTGGGCTTCTGGGCTCAAGGAGAAACAGGAGCAACGGTTCCTCCAGGAGAACCCAG 2047
Db 1331 TTCCTGGAGCTCTGGACTGATGGAGCCCGGGTCTCTCCAGGACCAAGCTGGTCTAATG 1390
QY 2048 GATACATGGGTTTACCGGGATTCAGGAAACAAAGGGGCAAGGAAATCAAGGTGAA 2107
Db 1391 GTGCTCTGGACTCGAGGTGTGCAAGTGTGATGCTGTGAAGATGTGCAAGGAGAGC 1450
QY 2108 AAGGTATTCAGGGTCAAAAGGAGAAATGCAAGACAGGAAATTCAGGCAACAGGAA 2167
Db 1451 CCGACACAGTGGTGAACCGCGTGAAGCTGTATTCAGGTGTTCAGGAGCTAAGCGG 1510
QY 2168 TTCAGGCAATCATGGTGCAAAAGAGAGAGAGGTGAAAGGGAGAACCTGGTGTCCGAG 2227
Db 1511 AAGATGCAAGGATGGATCACTGGAGAACCTGTGCAAAATGGGCTTCCAGGAGCTGCAG 1570
QY 2228 GTGCCATTGGATCAAAAGGAGAAATCTGGGTGTGATGCTGTGAGGCGCCGAGGTCCTA 2287
Db 1571 GAGAAAGGGTGCCCTCGGTTCGAGACCTGTGCAACAAATGGCATCCAGGAGAA 1630
QY 2288 AGGGGCAACCTGGGATCCAGGTCTCAGGACCCCGAGGTTTGGATGGGAAACCCGGA 2347
Db 1631 AGGTCTCTGGAGAGCGTGTCTCCAGGCCCTGCAGGCGCCAGAGGAGCTGCTGGAG 1690
QY 2348 GAGAGTTTTCAGAACATTTATTCAGACAGTTTGCACAGATGTATTAAGAGCCAGCTAC 2407
Db 1691 AACCTGGCAGAGATGGGTCCTCGAGGTCCAGGAATGAGGGGATGC---CCGGAATC 1747
QY 2408 CAGTCTTACTTCAGAGTGGAAAGATTAAGAAATTTGTGATCATTTGCTGTCCCAACATGGCT 2467
Db 1748 CAGGAGNACAGGAGTGTGGAACCCAGGCTCCCGGAGTCAAGGAGAAAGTGTGTC 1807
QY 2468 CCCCGGATTCCTGGGCACTCTGTTCGATAGGCCAGAGGTTCCAGAGGATTAACCTG 2527
Db 1808 GACCAGGTCCTCTGGGCACTCTGTCTCCGAGGTGAGCTGTGTATGGGCTTCCCG 1867
QY 2528 GTTTCAGGAGAGATGGTGTCTCTGATTAAGTGGGTGCTCTGGAGCTCCAGGTGTCA 2587
Db 1868 GTCTTAAGGAAATGATGCTCTCTGTAAAGATGAGAACAGGTGGCCCTTGAGGAC 1927
QY 2588 GAGGATTAAGAGGCTACCAAGAGAAATGGGAAAGAGGAGCCAAAGGTTTGGTATC 2647
Db 1928 CTGGCCCTCAGGCTCTCTGGAAGAAATGGTGAACCTGGACCTCAAGGACCCCGAGGC 1987
QY 2648 CTGGAGAACAGGCTCTCTGTCCTCCAGGTCCAGGCGCCCTCTGGAATAGCAAG 2707
Db 1988 CTACTGGCCCTGG---TGGTGAAGAGAGACACAGGACCCCTCTGTCACAGGATTAC 2044
QY 2708 AAGGTCTCCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGACCATCGAAACCTGGAA 2767
Db 2045 AAGGCTTCCCTGGTACAGTGTCTCTCCAGGAGAAATGGAAACCTGGGGAACCAAGTC 2104
QY 2768 TCCAGGCAACAGGCCCCCAGG 2792
Db 2105 CAAGGGTGTATGCGGTGCACCTGG 2129

```

RESULT 13
 US-10-096-534-15
 ; Sequence 15, Application US/10096534
 ; Publication No. US2003016887A1
 ; GENERAL INFORMATION:
 ; APPLICANT: The Brigham and Women's Hospital, Inc.
 ; APPLICANT: Yatez, Karen
 ; APPLICANT: Mizuno, Shuichi
 ; APPLICANT: Glowacki, Julie
 ; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF SKELETAL DEGENERATION CONDITIONS
 ; FILE REFERENCE: B0801/7244/KA/ERP

```

; CURRENT APPLICATION NUMBER: US/10/096,534
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: JS 60/274,980
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 6319
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-096-534-15

Query Match: 8.4%; Score 241.4; DB 12; Length 6319;
Best Local Similarity 50.9%; Pred. No. 2e-55;
Matches 757; Conservative 0; Mismatches 666; Indels 4; Gaps 6;

QY 1326 TTGTCCTCCGGAAAAACAGGACTTCAAGSCCCCAAGGTGACCTGGACTSCTGGAA 1385
Db 3423 TGGAGCTCAGGTGCACCTGGACTGNAAGAGAGGGAAGGTCCCGAGGCCACAGGTCC 3482
QY 1386 CCCTGGCTACCTCGAACCAACCTGTCAGATGTAAGCCTGSAATATCAGGGAATTGCGG 1445
Db 3483 AGTTGGCTCACAGAGAACCTGGGTGAGCAGGTACAGCTGGCCCAATTGGTTACAGG 3542
QY 1446 GACACCAAGTGTTCAGGATCTCCAGGAATACAGGAGCTCGAGGACTACAGGTTCACAA 1505
Db 3543 GCGCCCGGACCTCAGGCTCTCTGCTCCAGCTGGAGAGAAAGGTCTCTCGAGAGAAA 3602
QY 1506 AGGAGAACCGGGCGAGATGTGACAGGGTGA*CGTGGACTTCTCTGTTTCTCTGGGCT 1565
Db 3603 AGTCCCCAAGGGCTTCGAGGAGAGATGGAGTTCAGGTCTCTGTTCTCCAGGGCC 3662
QY 1566 TCATGGCATGCCAGGATCAAAAGGTGAATGGTSCCAAGAGAGACAAAGGATCACCTGG 1625
Db 3663 AGCTGTCTCTCGGCTCCCTGGGAAAGCGAGACAAGGTGAAATGGTGAAGCCGG 3722
QY 1626 ATTTATGGCAAAAGGTTGCAAGGTGAAAGGGGAATGCTGGCTTCCCTGGCTCCC 1685
Db 3723 ACNAAAGGAGCAGGAGGTGACAAGGGAGAAATGGCCCTCCCGTCCCGCAGGTCTCA 3782
QY 1696 TGGACCTCTCGAGAACCAAGGACATGGAAGGATGGATTAATGGGTAGTCCCGGTTT 1745
Db 3783 AGGACCACTGTGTCCTCTGGAATTCCTGGAGGTGATGTCNACAGGTCTTAGAGACA 3842
QY 1746 CAAGGAGAACAGGATCCCTGCTGTCGCGGCGAGGATGGAAACACGCGGAGAGCCCTGG 1805
Db 3843 GCAGGGGATGTTTGGGCAAAAGGTGATGAGGTGSCCAGAGGCTTCCCTGGACCTCTGG 3902
QY 1806 AATCCCAAGATTTCTCTGAAACCCAGGATTAATGGGCCAAAGGGAGAAATTT----- 1857
Db 3903 TCCAATAGTCTTCAGGCTGTCGCCGCCNCCCTGGTGAAGGATGAATGGGGATGT 3962
QY 1858 -----GGGCTTCAGACAGCAAGGAAAAAGAGCCCGAGGATGCTGGTTT 1907
Db 3963 TGTTCATGGGGCCACCTGCTCTCCAGGCCCAAGAGGCGCTCAAGGTCCCAATGGAGC 4022
QY 1908 ATGGGAAGCAATGCTCACAGGCCAGCTCGAACACCCGGATCTAAGGAGCAAGG 1967
Db 4023 TGATGGACCAAGGACCCCGAGGTTCGTGTGTT*CACTTGGTGTGTTGGAGAAAGGG 4082
QY 1968 TGAACCTGGAATTCAGGGGATGCTCGGGCTTCTGGGCTCAAGGGAGAACAGAGCAAC 2027
Db 4083 TGAACCTGGAAGCAAGGAAACCCAGGGCTCTGGGGAGGAGGTGTAGCGGTCCCAA 4142
QY 2028 GGGTTCCTCCAGGAGAACCCAGGATACATGGTTCCTGTTTACCCTGGGATTCAGGAAAAAGGGGA 2087
Db 4143 AGGAGAAAGAGAGAGAAAGGGGAAGGTGCTCCACCTGGA3CTGCTGGAGCTCCAGGTGC 4202
QY 2088 CAAAGGAATCAAGGTGAAAGGATATTCAGGTCAAAAGGGAGAGAAATGGMAGACAGGG 2147
Db 4203 CAAGGGCGCCAGGTGATGATGGCCCTTAAGGCCCTAAGCGGTACCCGGGCTCTGTTCTCTGG 4262
QY 2148 AATTCAGGGCAACAGGGAA*TCAGGCCATCATGGT*GCAAAAGGAGAGAGAGGTGAAA 2207

```

1688 GACCTGCTGGAGACCCAGGAGACATGAAAGGATGGATTAAATGGGTAGTCCCGTTTCA 1747
 1025 GGGCTCTCTGGTGAGCGAGAGCCAGGACTCTCTGGGGCTGCAGGTGCTCGGGGTAAATG 1084
 1748 AGGAGAAAGCAGGATCCCTCTGGTCTCCGGGCGAGATGGAAACACGGGAGAGCTCGAA 1807
 1085 ACGGTGCTCGAGCGAGTATGTCACACAGGCCCTCTGGTCTCTCTGGAACTCCGGAT 1144
 1808 TCCAGAGATTTCTCGAAACGAGATTAATGGGCCAAAAGGAGAAATGGGCTCCAG 1867
 1145 TCCCTGGATCCCTGGTCTAAGGGTGAATGGACCTCCAGGGTCTCTGGTTCAATG 1204
 1868 GACAGCAAGAAAAAAGGAGCCCGAGGATCCCTGGTTTAAATGGGAAGCAATGGCTCAC 1927
 1205 GTGCCCTGGACAAAGAGAGAACTCTGGACCTCAGGACACGCTGGTCTCAAGTCTTC 1264
 1928 CAGGCAGCTCGAAACACCGGATCTAAGGGAGCAAAAGGTGAACCTTGAATTCAGGGA 1987
 1265 CTGGCCCTCTGGGATTAATGTAGTCTCTGGTGTAAAGGGGAAATGGGTCTCCGCTGCA 1324
 1988 TGCCTGGGCTTCTGGGCTCAAGGAGCAACACGAGGCAACGGTTCCTCCAGGAAACGAG 2047
 1325 TTCTTGGAGCTCTGGAGTGTGAGGCCCGGGTCTCTCAGGACACGCTGGTCTAATG 1384
 2048 GATACATGGGTTTACCCGGGATTTCAAGGAAAAAAGGGGCAAAAGGAAATCAAGGTGAAA 2107
 1385 GTGCTCTGGAGCTGGAGGTGGTGCAGGTGAGCTGTGTAAGATGGTGCCTCAAGGAGC 1444
 2108 AAGGTATTTCAGGCTCAAAAGGAGAAATGGGAGACAGGGAATCCAGGCAACAGGAA 2167
 1445 CCGGACACGCTGGTGAACCGGTGAGGTGGTATTTCAGGTGTTCCAGGAGCTAAGGCG 1504
 2168 TTCAAGGCGCATCATGGTGCMAAGGAGAGAGAGGTGAAGAGGAGAACCTGGTTCGCGAG 2227
 1505 AAGATGGCAAGGATGGATCACTCTGGAGAACCTGTGTGCAAAGGGCTTCAGGAGTGCAG 1564
 2228 GTGCCATTTGGATCAAAAGGAAATCTGGGGTGGATGGCTTGAATGGGGCCGAGAGTCTTA 2287
 1565 GAGAAAGGGGTGCCCCCTGGGTTCCGAGGACCTGTGTGACCAAAATGGCATCCGAGAGAAA 1624
 2288 AGGGCAACCTGGGATCCAGGTCTCAGGACCCCGAGGTTTGGATGGGAAGCCCGGAA 2347
 1625 AAGGTCTCTGAGAGCTGGTGTCTCAGGCCCTTCAGGGCCCGAGAGAGCTGTGGAG 1684
 2348 GAGAGTTTTCAGAACAAATTTATTCGACAAATTTTGACAGATGTATTAAGAGCCAGCTAC 2407
 1685 AACCTGGCAGAGATGGCTCCCTGGAGGTCCAGGAATGAGGGCATGC---CCGGAAGTC 1741
 2408 CAGTCTTACTTCAGAGTGAAGAAATAGAAATTTGTGATCATTTGCTGTCCCAACATGGCT 2467
 1742 CAGGAGGACAGAGATGTGGAAACACAGGCGCTCCCGAAGTCAAGAGAAATGGTGTGTC 1801
 2468 CCGGGTATTCTGGGCCACCTGGTCCGATAGGCCCCAGAGGGTCCGAGAGATTTACCTG 2527
 1802 GAGCAGCTCTCTGGGCCATCTGGTCCCGAGGTGAGCTGTGTCTGATGGCTTCCCGG 1861
 2528 GTTTCAGAGAGAGATGTGTCTCTGGATTTAGTGGGTCTCCTGGAGCTCCAGGTGTCA 2587
 1862 GTCTTAAGGAAATATGTGTCTCTGGTGAAGATGGAGAACGAGGTGGCCCTGGAGAAC 1921
 2588 GAGGATTAAGAGCCTACAGGAAGAAATGGGGAAGAAAGGGAGCAAGGTTTGGGTATC 2647
 1922 CTGGCCCTCAGGCTCTCTCTGGAAAGATGGTGAATCTGGAACCTCAAGGACCCCGAGG 1981
 2648 CTGGAAACAGGTCTCTCTGGTTCCTCCAGGTCCTCAGAGGGCCCTCTCTGGAATAGCAAG 2707
 1982 CTACTGGGCTGG---TGGTGAACAAAGGAGACACAGGACCCCTGTGTCTCACAAGATTAC 2038
 2708 AAGGTCTTCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGACCATGGAAACCTGGAA 2767
 2039 AAGGCTTGGCTGTACAGGTGGTCTCTCAGGAGAAAAATGGAAACCTGGGAAACAGGTC 2098

2768 TCCAAGGGCAACACGAGCCCCCAGG 2792
 2099 CAAAGGGTGTGCCGGTGCACCTGG 2123

RESULT 12
 US-10-198-646-9949
 ; Sequence 9949, Application JS(10-98846
 ; Publication No. US2003009974A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lillie, James
 ; APPLICANT: Xu, Yongyao
 ; APPLICANT: Wang, Youzhen
 ; APPLICANT: Steinmann, Kathleen
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
 ; FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
 ; THERAPY OF BREAST CANCER
 ; FILE REFERENCE: MRI-049
 ; CURRENT APPLICATION NUMBER: US/10-198,846
 ; PRIOR FILING DATE: 2002-07-18
 ; PRIOR APPLICATION NUMBER: 60/306,220
 ; NUMBER OF SEQ ID NOS: 14084
 ; SOFTWARE: PastSeq for Windows Version 4.0
 ; SEQ ID NO 9949
 ; LENGTH: 8664
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-198-646-9949

Query Match: 8.5%; Score 242.6; DB 14; Length 8564;
 Best Local Similarity 48.8%; Pred. No. 1.2e-55;
 Matches 715; Conservative 0; Mismatches 744; Indels 6; Gaps 2;

1328 GTCCCTCCGGAAGAACAGGACTTCAAGGCCCAAGAGTGCACCTGGACTGCTGGGAACC 1387
 671 GTCCCTCCGGAAGAACAGGACTTCAAGGCCCAAGAGTGCACCTGGACTGCTGGGAACC 730
 1388 CTGGTACCTCGAACAACCTGTCAAGATGTGAAGCTGGATATCAGGGAATTCAGGGA 1447
 731 CTGTGAACCTGGGAAGCTGTCTCTCAGGCCCCCAGGACCTCTGTGTATAGTGC 790
 1448 CACGAGTGTTCAGGATCTCCAGGAATACAGGAGCTCGAGGACTACAGGTACAAAG 1507
 791 CATCTGTCTCTGGAAGATGGAGATCAGGTAGACCCGACGACTGGAGAGCAG 850
 1508 GAGAACACAGGCGGAGATGGTGAAGGATGATCGTGACTTCTGTGTTTCTCTGGCTTC 1567
 851 GATTGCTGGACCTCCAGGTA*CAAAGGTCCAGCTGGGATACCTGGATTCCTGGTATGA 910
 1568 ATGGCATGCCAGATCAAGGGTGAATGGGTGCCAAGGAGACAAAGATCACCTGGAT 1627
 911 AAGGACACAGAGGCTTCGATGACGAATGGAGAAAGGTAAGGTAAGGTAAGGTAAGG 970
 1628 TTTATGGCAAAAGGCTGCAAAAGGTGAAAAGGGGAATGCTGGCTTCCTGTGGCTTCCCTG 1687
 971 TAAAGGCTGAATGGTCTTCCAGGGAATGGAGCTCTCGACCCATGGTCCCAAGAG 1039
 1688 GACTCTCTGGAGAACACAGGAACATGGAAGGATGGATTAATGGGTAGTCCCGTTTCA 1747
 1031 GGGCTCTGTGTGAGCAGACCGCCAGGACTTCTCTGGGGCTGCAGGCTCTCGGGTAATG 1090
 1748 AGGGAAGACAGATCCCTGTGTCTCCGGGCGAGGATGGAACACGGGAGAGCCTGGAA 1807
 1091 ACGGTCTCGAGCAG*GATGGTCAACAGGCCCTCTCTGTGCTCTCTGAACTGCCGAT 1150
 1808 TCCAGGATTTCTCGGAACCGAGGATTAATGGGCCAAAGGAGAAATGGGCTTCCAG 1867
 1151 TCCCTGGATCCCTGTGTCTAAGGGTGAATGGACCTGCAGGGTCTCTCTGTCAAATG 1210
 1868 GACAGCAAGAAAAAAGGAGCCCGAGGATGCCTGTGTTTAAATGGGAAGCAATGGCTCAC 1927
 1211 GTGCCCTCGCAAAAGAGGAGAACCTGGACCTCAGGAGACACGCTGTGTCTCAAGGTCTCTC 1270

```
Db 710 CTGGTGAACCTGGGCAAGCTGGTCTCTTCAAGGCCCTCCAGGACCTCTCTGGTGTATAGGTC 769
Qy 1448 CACCAGGTTTCCAGGATCTCCAGGAATACAGAGAGCTCCAGGACTACCAAGTTACAAG 1507
Db 770 CATCTGGTCTCTGGAAAGATGGAGATCAGTATAGACCCGACGACCTTGGAGAGCGAG 829
Qy 1508 GAGAACCAAGGCGAGATGGTCAAGGGTGTCTGGACTTCTCTGGTTTCTCTGGGCTTC 1567
Db 830 GATTGCTGGACCTCCAGGTATCAAGGTCCAGCTGGATACCTGGATTCTCTGGTATGA 889
Qy 1568 ATGGCATGCCAGATCAAGGGTGAATGGTGCACAAAGAGACAAGGATCACTGGAT 1627
Db 890 AAGGACACAGAGGCTTCGATGGACGAATGGAGAAAGGGTGAACAGAGTCTCTCTGGAT 949
Qy 1628 TTATATGCAAAAGGGTGCACAAAGGTGAAGAGGGAATGTCTGGCTTCTCTGGCTCCCTG 1687
Db 950 TAAAGGTGAATGGTCTTCCAGCGCAATATGAGCTCTCTGACCATGGTCCCAAGAG 1009
Qy 1688 GACCTGTGGAGAACCAAGGAAGACATGGAAGAGATGATTAATGGGTAGTCCCGGTTTCA 1747
Db 1010 GGGCTCTCTGGTGAAGAGCGGCGAGACTTCTCTGGGGCTGACAGGTCTCTGGGGTAATG 1069
Qy 1748 AGGAGAGACAGGATCCCTCTGGTCTCTGGGGAGGATGGAAACAGGGGAGAGCTCGAA 1807
Db 1070 ACGGTGCTGAGGACGATGATGTCACAGGACCTCTCTGGTCTCTCTGGAACTCCCGGAT 1129
Qy 1808 TCCAGGATTTCTGGAAACCGAGATTAATGGGCCAAAGAGGAGAAATTTGGGCTCCAG 1867
Db 1130 TCCCTGGATCCCTGGTCTAAGGGTGAAGTTGACCTCGCAGGGTCTCTGGTTCAATG 1189
Qy 1868 GACAGCAAGGAAGAAAGAGGACCCAGGATGCTCTGGTTTAAATGGGAAGCAATGGCTCAC 1927
Db 1190 GTGCCCTCTGGCAAGAGAGAACTCGACCTCAGGACACAGCTGGTGGCTCAAGGTCCTC 1249
Qy 1928 CAGGCGACCTGGAAACCGGGATCTAAGGGAAGCAAGGTGAACCTTGGAAATCAAGGGA 1987
Db 1250 CTGGCCCTCTGGGATTAATGGTAGTCTCTGGTGAAGCGAATGGTCCCGCTGGCA 1309
Qy 1988 TGCTCTGGGCTTCTGGGTCAAGGGAGAACAGGAGCAACCGGTTCCTCCAGGAGAACAG 2047
Db 1310 TTCTCTGAGCTCTCTGGATGGAGCGCGGGTCTCTCCAGGACCAAGCGGTCTAATG 1369
Qy 2048 GATACATGGTTTACCCGGATTCAGGAAAGAGGGAAGAGGGAAGAAATCAAGGTGAAA 2107
Db 1370 GTGCTCTGGACTCGAGGTGGTGCAGGTGAGCTGGTGAAGATGGTGCACAAAGGAGGC 1429
Qy 2108 AAGGTATTTCAGGTCACAAAGGGAGAGAAATGGAAGACAGGGAATTCAGGGCAACAGGGA 2167
Db 1430 CCGGACCAAGTGGTGAACCGGTGAGGCTGGTATTCCAGGTGTTCCAGGAGCTAAGGCG 1489
Qy 2168 TTCAAGGCTCATGGTGCAGAAAGAGAGAGAGGTGAAGAGGAGAACCTGGTCTCCAG 2227
Db 1490 AAGATGGCAAGATGGATCACTTGAGAACCTGGTGCACAAATGGGCTTCCAGGAGCTCAG 1549
Qy 2228 GTGCCATTGGATCAAAAGAGAAATCTGGGGTGGATGGCTTCACTGGGGCCCGAGTCTTA 2287
Db 1550 GAGAAAGGGTGCCTCTGGTTCGAGAGACCTGTGGACCAATGGATCCAGAGAGAA 1609
Qy 2288 AGGGCAACCTGGGATCCAGGTCTCTAGGACCCCGGAGTTTGGATGGGAAGCCCGGAA 2347
Db 1610 AGGGTCTCTGGAGAGGCTGGTGTCTCCAGGCCCTGCGAGGCCCGCAGGAGGAGCTCTGGAG 1669
Qy 2348 GAGAGTTTTCAGAAATTTATTCACAAAGTTTGCACAGATGTAATAGAGCCCGAGCTAC 2407
Db 1670 AACCTGGCAGAGATGGGCTCTCTGGAGTCCAGGAATCAGGGGCGATGC---CCGGAAGTC 1726
Qy 2408 CAGTCTTACTTTCAGAGTGAAGAAATAGAAATTTGTGATCATTTGCTGTCTCCCAACATGGCT 2467
Db 1727 CAGGAGGACCAAGAGATGATGGAAACCAAGGCCCTCCCGGAGTCAAGGAGAAATGGTCTC 1786
Qy 2468 CCGCGGATTTCTCTGGGCACTGTGTCCGATAGGCCCCAGAGGGTCCCAAGAGGATTACCTG 2527
```

```
Db 1787 GACCAGGTCTCTCTGGGCCATCTGGTCCCCGAGGTCAAGCTGGTGTCTATGGGTTCCCCG 1846
Qy 2528 GTTTCGCAAGGAAGATGGTGTCTCTGGATTAGTGGGTGTCTCTGGACGTCCAGGTGTCA 2587
Db 1847 GTCTAAAGGAATGATGGTGTCTCTGGTAAAGTGGAGACGAGGTGGCCCTGGAGGAC 1906
Qy 2588 GAGGATTAAGGCTTACCAGGAAGAAATGGGAAAAAGGGAGCCAAAGGGTTTGGGTATC 2647
Db 1907 CTGGCCCTCAGGGTCTCTCTGGAAAAAGATGTGAACTGGAACTCAAGGACCCCGAGGC 1966
Qy 2648 CTGGAGAACAGGTCTCTCTCTGGTCCCCCAGGTCCAGAGGGCCCTCTGGAAATAGCAAG 2707
Db 1967 CTACTGGGCCCTGG---TGGTGACAAAGAGACACAGGACCCCTCTGGTCCCAAGGATTAC 2023
Qy 2708 AAGGTCTCTCCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGACCATGGAAAACTGGAA 2767
Db 2024 AAGGCTTCTCTGGTACAGGTGTCTCTCCAGGAGAAATGGAAAACTGGGGAACCAAGTC 2083
Qy 2768 TCCAAGGCAACCAAGGCCCCCGG 2792
Db 2084 CAAAGGATGATGCCCGTGCACCTG 2108

RESULT 11
US-10-096-534-13
; Sequence 13, Application US/10096534
; Publication No. US2003016687A1
; GENERAL INFORMATION:
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Yates, Karen
; APPLICANT: Mizuno, Shuichi
; APPLICANT: G.owacki, Julie
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF SKELETAL DEGENERATION CONDITIONS
; FILE REFERENCE: B08017244/KA/ERP
; CURRENT APPLICATION NUMBER: US/10/096,534
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: US 60/274,980
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 5489
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-096-534-13

Query Watch 8.5%; Score 242.6; DB 12; Length 5489;
Best Local Similarity 48.8%; Pred. No. 8.3e-56;
Matches 725; Conservative 0; Mismatches 744; Indels 6; Gaps 2;

Qy 1328 GTCTCTCCGGAAACAGGACTTCAAGGCCCCCAAGGTGACCTGGAGTGGCTCGGAACC 1387
Db 665 GTCCCTCTGGTACATCTGGTCACTCTGGTTCCTGGATCTCCAGGATACCAAGGACCCC 724
Qy 1389 CTGGCTACCTCGGACCAAGCTTGAAGATGGTAAGCTTGGATATCAGGGAATTCAGGGA 1447
Db 725 CTGGTAACCTGGGCAAGCTGGTCTTCAGGCCCTCCAGGACCTCTGGTCTATAGTTC 784
Qy 1448 CACCAAGGTGTTCCAGGATCTCCAGGAATACAGGAGCTTCAGGAGTACCAAGTACAAAG 1507
Db 785 CATCTGGTCTCTGGAAAAAGATGAGAAATCAGGTAGACCCCGGACGACCTGGAGAGCGAG 844
Qy 1508 GAGAACCAAGGCGAGATGGTGCACAGGATGATCGTGGACTCTCTGGTTTCTCTGGGCTTC 1567
Db 845 GATTGCTTGGACCTCCAGGTATCAAGGTCCAGCTGGGATACCTGGATTCTCTGGTATGA 904
Qy 1569 ATGGCATGCCAGGATCAAGGGTGAATGGGTGCCAAGGAGACAAAGGATCACTGGAT 1627
Db 909 AAGGACACAGAGGCTTCGATGGAGAAATGGAGAAAGGGTGAACAGAGTGTCTCTGGAT 964
Qy 1628 TTATATGCAAAAGGGTGCACAAAGGTGAAGAGGGAATGTGGGTTCCTGGCTCTCTG 1687
Db 965 TAAAGGTTGAAATGGTCTCTCCAGCGGAAAAATGGAGTCTCTGGACCCATGGGTCCAAAGAG 1024
```

QY 1508 GAGAACCCAGGCGAGATGGTGACAAGGGTGATCGTGGACTTCTCGTTTCTCTGGGCTTC 1567
 DB 830 GATTGCTCGACCTCCAGGTATCAAGGTCAGCTGGATACCTGGATTCCCTGGTATGA 889
 QY 1568 ATGGCATCCAGGATCAAGAGGGTGAATGGTGCAGAGAGACAAGGATCACTGGAT 1627
 DB 890 AAGGACACAGAGGCTTCGATGGAGCAATGGAGAAAGGGTGAAAGGGTGCTCTGGAT 949
 QY 1628 TTTATGCAAAAAGGGTGCAAAAAGGTCAAAAAGGGGAATGCTGGCTTCCCTGGGCTCCCTG 1687
 DB 950 TAAAGGGTGAATGGTCTCCAGGCGAATGGAGCTCTGACCAATGGTCCAGAG 1009
 QY 1688 GACTGCTGAGAACCAAGGAGACATGGAAAGAGATGAATTAATGGGTAGTCCCGGTTTCA 1747
 DB 1010 GGGCTCTGCTGAGCGAGGACGGCCAGCACTTCTGGGGCTGCAGGTGCTCGGGGTAATG 1069
 QY 1748 AGGAGAAAGCAGGATCCCTGCTGCTCCGGGAGGATGGAACACGGGAGAGCCCTGGAA 1807
 DB 1070 ACGTGTCTGAGGAGTATGGTCAACAGGCGCTCTCTGTCTCTCTGAACTGCCGAT 1129
 QY 1808 TCCAGGATTTCTGGAAACCGAGGATTAATGGGCCAAAAGGGAGAAATTTGGGCTCCAG 1867
 DB 1130 TCCCTGATCCCTGGTCTAAGGGTGAAGTTGGACCTGCAGGCTCTCTGGTTCAAATG 1189
 QY 1868 GACAGCAAGGAAAAGGAGCCCGGAGTCTGTTTAAATGGGAGCAATGGCTCAC 1927
 DB 1190 GTGCCCCTGACAAAGGAGAACTCGACTCAGGACACGCTGGTGTCTCAAGGTCTCT 1249
 QY 1928 CAGCCAGCCTGGAACCCGGATCTAAGGAGCAAAAGGTGAACCTGGAATCAAGGGA 1987
 DB 1250 CTGGCCCTCTGGATTAATGGTAGTCTGTGTAAAGGCGAATGGTCCGCTGGCA 1309
 QY 1988 TGCCTGGGCTTTGGGCTCAAGGAGAAACAGAGCAACCGGTTTCCCAAGGAGAACCCAG 2047
 DB 1310 TTCTGGAGCTCTGGACTATGGAGCCCGGGTCTCTCCAGGACCAAGCCGGTCTAATG 1369
 QY 2048 GATACATGGGTTTACCCGGATTCAGGAAAAGGGGCAAGAGGAATCAAGGTGAA 2107
 DB 1370 GTGCTCTGACTCGAGGTGCTGAGTGAGCTGTGAAGATGGTGCCTCAAGGAGAGC 1429
 QY 2108 AAGGTATTCAGGCTCAAAAGGGAGAAATGGAACAGGGAATTCAGGCGCAACAGGAA 2167
 DB 1430 CCGACACCGTGTGAACGGTCAAGCTGTGTATTCAGGTGTTCAGGAGCTAAGGCG 1485
 QY 2168 TTCAAGGCTATGTTGGTCAAAAGGAGAGAGAGTGAAAGAGGAGAACCTGTGTCCGAG 2227
 DB 1490 AAGATGGCAAGGATGATCACTCGAGAACCTGTGTGCAATGGGCTTTCCAGGAGCTG 1549
 QY 2228 GTGCCATTGGATCAAAAGGAGNACTCTGGGTGATGCTTGTATGGGCGCCGAGTCTTA 2287
 DB 1550 GAGAAAGGGTGCCCTGGGTTCGAGACCTGTGTGACCAATGGCATCCAGGAGAA 1609
 QY 2288 AGGGGCAACCTGGGATCCAGGTCTCAGGACCCCGAGGTTTGAATGGAAAGCCGGAA 2347
 DB 1610 AGGCTCTGTGGAGAGCGTGTCTCCAGGCGCTGCGAGGCGCCAGAGGAGCTGTGGAG 1669
 QY 2348 GAGAGTTTTCAGAACATTTATTCGACAGTTTTCACAGATGTAATAGAGCCAGCTAC 2407
 DB 1670 AACCTGCGAGAGTGGGCTCCCTGGAGGTCCAGGAATGAGGGGCGATGC---CCGAGTC 1726
 QY 2408 CAGTCTTACTTCAGAGTGAAGAAATAGAAATTTGTATCTGCTGTCCCAACATGGCT 2467
 DB 1727 CAGGAGGACAGGAGTATGGGAAACAGGGGCTCCCGGAGTCAAGGAGAAATGGTCT 1786
 QY 2468 CCCGGGATTTCTGGGCCACTGTGTCCGATAGGCCCGAGAGGGTCCCGAGGATTAACCTG 2527
 DB 1787 GACCAGGTCTCTCTGGGCACTGTGTCCCGAGGTGAGCCCTGTGTCTATGGCTTCCCG 1846
 QY 2528 GTTTCAGGAGAGATGCTGTCTCTGTATGATGTTGTTGCTCTCTGGAGTCCAGGTGCA 2587
 DB 1847 GTCTTAAAGGAAATGATGGTCTCTCTGTAAAGATGGAGAACAGAGGTGGGCTGGAGAC 1906
 QY 2588 GAGGATTAAGAGGCTTACCAGAGAAATGGGAAAAGGGAGCAAGGGTTTGGGTATC 2647

DB 1907 CTGGCCCTCAGGCTCTCTCTGAAAGAAATGGTGAATCTGGACCTCAAGGACCCCAAGGC 1966
 QY 2648 CTGGAGAACAGGCTCTCTCTGTCCTCCAGGTCCAGAGGCGCTCTCTGGATTAAGCAAG 2707
 DB 1967 CTACTGGGCTTGG---TGGTGACAAAGAGACACAGGACCCCTCGTCCACAGGATTAC 2023
 QY 2708 AAGGTCTCTCCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGACCATGGAACCTCGAA 2767
 DB 2024 AAGGCTCTGCTGTACAGGTGCTCTCCAGGAGAAATGGAATAACCTGGGAGACCAAGTC 2083
 QY 2768 TCCAAGGCAACAGGCGCCCGCAGG 2792
 DB 2084 CAAAGGCTGATGCCCGTSCACCTGG 2108

6

RESULT 10
 US-10-177-293-67
 ; Sequence 87, Application US/10:77293
 ; Publication No. US20030124128A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Glatt, James
 ; APPLICANT: Lillie, Karen
 ; APPLICANT: Zhao, Xumei
 ; APPLICANT: Gannavarpu, Manjula
 ; APPLICANT: Kamatkar, Shubhangi
 ; APPLICANT: Mertens, Maureen
 ; APPLICANT: Myer, Vic
 ; APPLICANT: Wang, Youzhen
 ; APPLICANT: Xu, Yongyao
 ; APPLICANT: Hoersch, Sebastian
 ; APPLICANT: Monahan, John
 ; APPLICANT: Meyers, Rachel E.
 ; APPLICANT: Bast Jr., Robert C.
 ; APPLICANT: Hortobagyi, Gabriel N.
 ; APPLICANT: Pusztai, Lajos
 ; APPLICANT: Meric, Funda
 ; APPLICANT: Sahin, Aysegul
 ; APPLICANT: Mills, Gordon B.
 ; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
 ; FILE REFERENCE: MR*-038
 ; PREVENTION, AND THERAPY OF BREAST CANCER
 ; CURRENT FILING DATE: 2002-06-21
 ; CURRENT APPLICATION NUMBER: US/10/177,293
 ; PRIOR APPLICATION NUMBER: US 60/299,887
 ; PRIOR FILING DATE: 2001-06-21
 ; PRIOR APPLICATION NUMBER: US 60/301,572
 ; PRIOR FILING DATE: 2001-06-27
 ; PRIOR APPLICATION NUMBER: US 60/306,501
 ; PRIOR FILING DATE: 2001-07-18
 ; PRIOR APPLICATION NUMBER: US 60/325,002
 ; PRIOR FILING DATE: 2001-09-25
 ; PRIOR APPLICATION NUMBER: US 60/362,585
 ; PRIOR FILING DATE: 2002-03-05
 ; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
 ; PRIOR FILING DATE: 2002-05-14
 ; NUMBER OF SEQ ID NOS: 506
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 67
 ; LENGTH: 5460
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-177-293-67

Query Match 8.5%; Score 242.6; DB 14; Length 5460;
 Best Local Similarity 48.8%; Pred. No. 8.3e-56;
 Matches 715; Conservative 0; Mismatches 744; Indels 6; Gaps 2;

QY 1328 GTCTCCCGGAAAACAGGACTTCAAGGCCCAAGGTGACCTGGACTGCTGGGAACC 1387
 DB 650 GTCCCCCTGGTACATCTGGTCTATCTGTTCTCCCTGGATCTCCAGGATACCAAGGACCCC 709
 QY 1388 CTGGCTACCTGGACAACCTGCTCAAGATGTTAAGCCTGGATATCAGGAAATTCAGGGA 1447

1446 GACACCAGGTGTTCCAGGATCTCCAGGAATACAAGGAGCTCGAGGACTACCAAGTTCACAA 1505
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3513 CGGCCGGGACCTCAGGGTCTCTCTGTGCCAGCTCGAGAGAAAAGGTCTCTCTGGAGAAAA 3572

1506 AGGAGAACCAAGGCAGATGCTGACNAGGGTGATCGTGGACTTCCTGGTTTTCTCTGGGCT 1565
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3573 AGGTCCCAGAAGGGCTCGAGGGAGAGATTGGAGTTCAAAGTCTGTGGTCTCCCAAGGGCC 3632

1566 TCATGGCATGCCAGGATCAAAAGGGTGAATGGGTGCCAAAGGAGAGACAAGAGATCACTGG 1625
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3633 AGCTGGTCTCGGGCTCCCTGGGAGAGACGAGACAAGGGTGAATTTGTTGAGCGGG 3692

1626 ATTTTATGGCAAAAAGGTGCAAAAGGTGAAAAAGGGAATGCTGGCTTCCTCTGGCTCC 1685
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3693 ACAAAAAGCGACNAGGGTGACAAGGAGAGAAATGGCCCTCCCGTCCCCCAGGTCTTCA 3752

1686 TGGACCTCTGAGNAACCAAGNAGACATGGAANGAT-----GGATT 1727
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3753 AGGACAGTTGGTCCCTCGAAATTGCTGGAGGTGATGGTGAACACAGTCTTAGAGGACA 3812

1728 AATGGGTAGTCCCGGTTTCAAGGAGAGCAGGATCCCTGGTCTCCGGGGCAGGATGG 1787
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3813 GCAGGGATGTTTGGCAAAAAGGTGATAGGGTGCCAGAGGCTTCCCTGACCTCTGG 3872

1788 AACCGGGGAGAGCTCGAATCCCAAGGATTTCTTGGAAACCGAGGATTAATGGGCCAAAA 1847
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3873 TCCAATAGTGTTTCAGGGTCTGCCAGGCCCACTGGTGA AAAAGGTGAAAATGGGGATGT 3932

1848 GGGAGAAATGGGCTCCAGACACAGCAAGGAAAAAAAAGGACCCAGGATGCTCGTTTT 1907
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3933 TGGTCCATGGGCCACCTGGTCTCAGGCCCAAGAGGCCCTCAAGGTCCCAATGGAGC 3992

1908 AATGGGAAGCAATGGCTCACAGGCCAGCTCGAACACCGGATCTAAGGGAAGCAAGG 1967
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3993 TGATGGACACAAGGACCCAGGTTCTGTGGTTTCAAGTTGGTGTGGAGAAAAGGG 4052

1968 TGAACCTGGAATTCAGGGATGCTGGGGCTTTGGGCTCAAGGGAGAACGAGGAGCAAC 2027
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4053 TGAACCTGGAGAAGCAGGAAAACCCAGGGCTCTCTGGGGAGACAGGTGTAGGCGGTGCCAA 4112

2028 GGGTTTCCCAGGAGAACAGGATACATGGGTTTACCGGGGATTCAGGAAAAAAGGSGGA 2087
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4113 AGGAGAAAGAGAGAAAAGGGAGCTGTCTCCTGGAGCTGCTGACCTCCAGGTGC 4172

2088 CAAAGGAAATCAAGGTGAAAAGGTATTACGGGTCAAAAGGGAGAAATGGAAGACAGGG 2147
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4173 CAAGGGGCCACAGGTGATGATGGCCCTCAAGGGTAAACCGGGTCTCTGTGGTTTTCTCGS 4232

2148 AATTCCAGGGCAACAGGAAATTCAGGGCCATCATGGTGC AAAAGGAGAGAGAGGTGAAA 2207
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4233 AGATLCTGTTCTCTTGGGAACTTGGCCCTGCAAGTCAAGATGGTGTGGTGGTGNCAA 4292

2208 GGGAGAACCTGTTGTCAGAGTGCCATTGGATCAAAAAGGAAATCTGGGGTGGATGSCITT 2267
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4293 GGGTCAAGATGAGATCTGTGTCAACCGGGTCTCTCGGCCCATCTGTGAGGCTGSCCC 4352

2268 GATGGGGCCGAGTCTTAAGGGCAACCTGGGGATCCAGGTCTCAGGGACCCCGAG 2327
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4353 ACCAGGTCTCTCTGGAAAACGAGGTCTCTCTGGAGCTGCAAGTGCAGAGGGAAGACAAGG 4412

2328 TTGTGATGGGAAGCCCGAAGAGAGTTTTTCAGACAATTTATTCGACAAGTTTGCACAG- 2386
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4413 TGAANAAGTGCTAAGGGGAGCAGGTGCAAGAGGTCCTCTCTGAAAAAACCGGCCCAGT 4472

2387 ATGTAAATAGACCCAGCTACAGTCTTACTTTCAGATGGGAAGAAATTAGAAATTGTGATC 2446
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4473 CGGTCTCTCAGGACCTGCAGGAAAGCTTGGTCCAGA-----AGGTCTTCGGGGC 4521

2447 ATTGCTGTCCCAACATGGCTCCCGGTTNTCTCGGGCCACCTGTGCTCCGNATAGGCCAG 2506
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4522 ATTCCTCTGTGTC-TGTGGAGAACAGGTCTCTCTGGAGCTGCAAGGCCAAGATAGGACAC 4579

QY	2507	AGGGTCCACAGAGGATTAACCTGGTTTCCAGGAGAGAGATGGTGTTCCTGGATTAAGTGGG	2566
DB	4580	CTGGTCTATGGGACCTCTCTGGCTTACCTGGTCTCAAGAGGTGACCTGGCTCCAAAGGGGTG	4639
QY	2567	TCCCTGCAGCTCCAGGTGTCCAGAGGATTAAAGAGCCCTACCAGGAAGAAATGGGGAAAAAG	2626
DB	4640	AAAAGGACATCTCTGGTTTAATTGGCTGATTTGGTCTCCAGAGAGACAGGGGAAAAAG	4699
QY	2627	GGAGCCAAAGGGTTTGGGTATCTCTGGAGAACAAAGTCTCTCTGGTCCCCCAGGTCCACAGG	2686
DB	4700	GTGACCCAGGGCTC-----CCTGGAACCTCAAGGATCTCCAGGAGCAAAAGGGATGGGG	4753
QY	2687	GCCTCTCTGSAATAAGCAAAAGAGTCTCCAGGAGACCCAGTCTCCCTGGCAAGATG	2746
DB	4754	GAATTCCTGGTCTCT-----GCTGGTCCCTTAGGTCCACTGGTCTCTCCAGGCTTACCAG	4807
QY	2747	GAGACATGAAACCTGGTAATCCAAGGGCAACAGGCCCCCCCCAGG	2793
DB	4808	GTCTGAGGCCCAAGGGTTAACAAAGGCTCTACTTGGACCCGTGGC	4854
RESULT 9			
US-10-301-822-32			
; Sequence 32, Application US/10301822			
; Publication: No. US20030148410A1			
; GENERAL INFORMATION:			
; APPLICANT: Millennium Pharmaceuticals, Inc.			
; APPLICANT: Berger, Allison			
; APPLICANT: Guillemette, Tracy L.			
; APPLICANT: Kamatkar, Shubhangi			
; APPLICANT: Schlegel, Robert			
; APPLICANT: Monahan, John E.			
; APPLICANT: Thibodeau, Stephen N.			
; APPLICANT: BURGART, Lawrence J.			
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND			
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND			
; TITLE OF INVENTION: THERAPY OF COLON CANCER			
; FILE REFERENCE: MEM01-029P2RM			
; CURRENT APPLICATION NUMBER: US/0301,822			
; CURRENT FILING DATE: 2002-11-21			
; PRIOR APPLICATION NUMBER: US 60/339,971			
; PRIOR FILING DATE: 2001-12-10			
; PRIOR APPLICATION NUMBER: US 60/361,978			
; PRIOR FILING DATE: 2002-03-05			
; PRIOR APPLICATION NUMBER: US 60/381,988			
; PRIOR FILING DATE: 2002-05-20			
; NUMBER OF SEQ ID NOS: 228			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 32			
; LENGTH: 5460			
; TYPE: DNA			
; ORGANISM: Homo Sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (103)...(4503)			
US-10-301-822-32			
Query Match			
Best Local Similarity 8.5%; Score 242.6; DB 12; Length 5460;			
Matches 715; Conservative 0; Mismatches 744; Indels 6; Gaps 2;			
QY	1328	GTCTCCGGGAAACACGAGACTTCAAGGCCCAAGAGTGACCTGGACTCCCTGGGAACC	1387
DB	650	GTCCCCCTGGTACATCTGGTCTATCTGGTTCCCTTGGATCTCCAGNTACCAAGGACCCC	709
QY	1388	CTGGCTACCTTGGAACAACCTGGTCAAGATGGTAAGCCTGGATATCATGGGMAATTCAGGGA	1447
DB	710	CTGGTGAACCTGGGCAAGCTGGTCTCTCAGGCCCTCCAGGACCTCTCTGGTGTATAGGTC	769
QY	1448	CACAGGTGTTCAGGATCTCCAGGAAATCAAGAGGCTCGAGGACTACCAAGTTACAAG	1507
DB	770	CATCTGTCTCTGTGGAAAAAGATGGGAATCAGGTAGACCCGGAACGACCTGGAGAGCGAG	829


```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-13890

Query Match      8.5%; Score 244.6; DB 14; Length 6545;
Best Local Similarity 51.0%; Pred. No. 2.6e-56;
Matches 759; Conservative 0; Mismatches 684; Indels 44; Gaps 6;

QY 1326 TTGCTCTCCGGGAAACACAGGACTTCAAGGCCCCCAAGGTGACCTCGACTCGCTGGGAA 1395
Db 1326 TTGCTCTCCGGGAAACACAGGACTTCAAGGCCCCCAAGGTGACCTCGACTCGCTGGGAA 1395
QY 3432 TGGAGCTCAGGTGACCTGACCTGAAAGGAGGAGGTCCCGAGGCCACACAGGTCC 3491
Db 3432 TGGAGCTCAGGTGACCTGACCTGAAAGGAGGAGGTCCCGAGGCCACACAGGTCC 3491
QY 1386 CCCTGGCTACCTGGACAACTGGTCAAGATGGTAAGCTTGATATCAGGGAATTCAGG 1445
Db 1386 CCCTGGCTACCTGGACAACTGGTCAAGATGGTAAGCTTGATATCAGGGAATTCAGG 1445
QY 3492 AGTTGGCTCACACAGGAAACGTGGTCAAGAGGTACAGCTGGCCCAATTTGTTTACCAGG 3551
Db 3492 AGTTGGCTCACACAGGAAACGTGGTCAAGAGGTACAGCTGGCCCAATTTGTTTACCAGG 3551
QY 1446 GACACACAGGTGTCAGGATCTCCAGGATACAGAGGAGCTCGAGGACTACAGGTACAA 1505
Db 1446 GACACACAGGTGTCAGGATCTCCAGGATACAGAGGAGCTCGAGGACTACAGGTACAA 1505
QY 3552 CGGCCCGGACCTCAGGCTGCTCTGGTCCAGCTGGAGAGAAAGTGTCTCTGGAGAA 3611
Db 3552 CGGCCCGGACCTCAGGCTGCTCTGGTCCAGCTGGAGAGAAAGTGTCTCTGGAGAA 3611
QY 1506 AGGAGAACACAGGCGAGATGGTGACAAAGGTGATCGTGGAATCTCTGGTCTTCTGGGCT 1565
Db 1506 AGGAGAACACAGGCGAGATGGTGACAAAGGTGATCGTGGAATCTCTGGTCTTCTGGGCT 1565
QY 3612 AGGTCCCAAGGCGCTGAGGAGAGATGGAGTTCAGGTCCTGTGTCTCCAGGCGC 3671
Db 3612 AGGTCCCAAGGCGCTGAGGAGAGATGGAGTTCAGGTCCTGTGTCTCCAGGCGC 3671
QY 1566 TCATGGCATGCGAGATCAAGAGGTGAAATGGTGCCAAAGGAGACAAAGATCACCTGG 1625
Db 1566 TCATGGCATGCGAGATCAAGAGGTGAAATGGTGCCAAAGGAGACAAAGATCACCTGG 1625
QY 3672 AGCTGGTCTGCGGCTCCCTGGGGAAGACGGAGACAAAGGTTGAAATTTGGTGAGCGCGG 3731
Db 3672 AGCTGGTCTGCGGCTCCCTGGGGAAGACGGAGACAAAGGTTGAAATTTGGTGAGCGCGG 3731
QY 1626 ATTTTATGGCAAAAGGTGCAAAAGGTGAAAGGGGAATGCTGGCTTCCCTGGCCTCCC 1685
Db 1626 ATTTTATGGCAAAAGGTGCAAAAGGTGAAAGGGGAATGCTGGCTTCCCTGGCCTCCC 1685
QY 3732 ACAAAGGAGCAAGGTGACAGGGAGAAATGGCCCTCCCGTCCCGCAGGTCTTCA 3791
Db 3732 ACAAAGGAGCAAGGTGACAGGGAGAAATGGCCCTCCCGTCCCGCAGGTCTTCA 3791
QY 1686 TGGACCTCTGAGAACACAGGAAGCATGGAAAGAT-----GGATT 1727
Db 1686 TGGACCTCTGAGAACACAGGAAGCATGGAAAGAT-----GGATT 1727
QY 3792 AGGACCAATTGGTGGCTGGAAATTTGCTGGAGGTGATGTTGAACACAGGTCTCTAGAGCA 3851
Db 3792 AGGACCAATTGGTGGCTGGAAATTTGCTGGAGGTGATGTTGAACACAGGTCTCTAGAGCA 3851
QY 1728 AATGGGTAGTCCCGTTTCAAGGAGAGACAGGATCCCTCTGGTCTCGGGGAGGATGG 1787
Db 1728 AATGGGTAGTCCCGTTTCAAGGAGAGACAGGATCCCTCTGGTCTCGGGGAGGATGG 1787
QY 3852 GCAGGGATGTTTGGGCAAAAGGTGATGAGGGTGCCAGAGGCTTCCCTGGACCTCTCG 3911
Db 3852 GCAGGGATGTTTGGGCAAAAGGTGATGAGGGTGCCAGAGGCTTCCCTGGACCTCTCG 3911
QY 1788 AACACGGGAGAGCTGGAAATCCCAAGGATTTCTGGAAACCGAGGATTAATGGGCCAAA 1847
Db 1788 AACACGGGAGAGCTGGAAATCCCAAGGATTTCTGGAAACCGAGGATTAATGGGCCAAA 1847
QY 3912 TCCATAGCTTTCAGGCTGCTCCAGGCCACCTGCTGTAAGAGGTGAAATGGGGATG 3971
Db 3912 TCCATAGCTTTCAGGCTGCTCCAGGCCACCTGCTGTAAGAGGTGAAATGGGGATG 3971
QY 1848 GGGAGAAATGGGCTCCAGGACAGCAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1907
Db 1848 GGGAGAAATGGGCTCCAGGACAGCAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1907
QY 3972 TGGTCCCATGGGCGACCTGGTCTCCAGGCCCAAGAGGCCCTCAAGGTCCCAATGGAGC 4031
Db 3972 TGGTCCCATGGGCGACCTGGTCTCCAGGCCCAAGAGGCCCTCAAGGTCCCAATGGAGC 4031
QY 1908 AATGGGAGCAATGCTCACAGGCCAGCTTGAACACCGGATCTAAGGAGAGCAAGG 1967
Db 1908 AATGGGAGCAATGCTCACAGGCCAGCTTGAACACCGGATCTAAGGAGAGCAAGG 1967
QY 4032 TGATGGACCAAGGAGACCCCGAGGCTGTTGGTTTCAGTTGGTGGTGTGGAGAAAGG 4091
Db 4032 TGATGGACCAAGGAGACCCCGAGGCTGTTGGTTTCAGTTGGTGGTGTGGAGAAAGG 4091
QY 1968 TGAACCTGGAAATCAAGGATGCTTGGGGCTTCTGGGCTCAAGGAGAGACCAAGGAGCAAC 2027
Db 1968 TGAACCTGGAAATCAAGGATGCTTGGGGCTTCTGGGCTCAAGGAGAGACCAAGGAGCAAC 2027
QY 4092 TGAACCTGGAGAGCAGGAGAACCCAGGCTTCTGGGAGAGAGGTGTAGCGGTCCCA 4151
Db 4092 TGAACCTGGAGAGCAGGAGAACCCAGGCTTCTGGGAGAGAGGTGTAGCGGTCCCA 4151
QY 2028 GGGTTCCCGAGAGAACCAAGATATCATGGTTTACCCCGGATTCAGGAGAAAGAGGGA 2087
Db 2028 GGGTTCCCGAGAGAACCAAGATATCATGGTTTACCCCGGATTCAGGAGAAAGAGGGA 2087
QY 4152 AGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4211
Db 4152 AGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4211
QY 2088 CAAAGGAAATCAAGGTGAAAGAGGTATTCAGGTCAAAAGGAGGAGAAATGGAAGCAGG 2147
Db 2088 CAAAGGAAATCAAGGTGAAAGAGGTATTCAGGTCAAAAGGAGGAGAAATGGAAGCAGG 2147
QY 4212 CAAAGGAGCAGAGTATGATGGCCCTAAGGTTAACCCGGGTCTGTGTGTTTCTCTGG 4271
Db 4212 CAAAGGAGCAGAGTATGATGGCCCTAAGGTTAACCCGGGTCTGTGTGTTTCTCTGG 4271
QY 2148 AATTCAGGAGCAAGGAAATCAAGGCCATCATGGTGCAGGAGAGAGAGAGAGAGAGAG 2207
Db 2148 AATTCAGGAGCAAGGAAATCAAGGCCATCATGGTGCAGGAGAGAGAGAGAGAGAGAG 2207
QY 4272 AGATCTGTCTCTCCGGGAGACCTGGCCCTGACAGTCAAGATGTTGTTGGTGTGCA 4331
Db 4272 AGATCTGTCTCTCCGGGAGACCTGGCCCTGACAGTCAAGATGTTGTTGGTGTGCA 4331
QY 2208 GGGAGAACTGTGTCTCGAGGTGCCATTTGGATCAAAAGGAGAAATCTGGGGTGGATGGCT 2267
Db 2208 GGGAGAACTGTGTCTCGAGGTGCCATTTGGATCAAAAGGAGAAATCTGGGGTGGATGGCT 2267
QY 4332 GGGTGAAGATGAGATTCCTGGTCAACCGGGTCTCTCTGGCCCATCTGTGTGAGGCTGGCC 4391
Db 4332 GGGTGAAGATGAGATTCCTGGTCAACCGGGTCTCTCTGGCCCATCTGTGTGAGGCTGGCC 4391
QY 2268 GATGGGGCCCGAGGTCTTAAAGGGCAACCTGGGATCCAGGTCTCTCAGGAGACCCCGAGG 2327
Db 2268 GATGGGGCCCGAGGTCTTAAAGGGCAACCTGGGATCCAGGTCTCTCAGGAGACCCCGAGG 2327
```

RESULT 8

```
US-10-084-817-155
; Sequence 155, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
; APPLICANT: Susan Stuart
; APPLICANT: Jed G. Nuchtern
; APPLICANT: Sharon E. Pion
; APPLICANT: Jason M. Shohet
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
; FILE REFERENCE: PA-0046 US
; CURRENT APPLICATION NUMBER: US/10/084,817
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 60/270,784
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 365
; SOFTWARE: PERL Program
; SEQ ID NO 155
; LENGTH: 6281
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030119009A1 3290944CB1
US-10-084-817-155
```

Query Match 8.5%; Score 243; DB 14; Length 6281;
Best Local Similarity 51.0%; Pred. No. 7.1e-56;
Matches 758; Conservative 0; Mismatches 685; Indels 44; Gaps 6;

```
QY 1326 TTGCTCTCCGGGAAACACAGGACTTCAAGGCCCCCAAGGTGACCTCGACTCGCTGGGAA 1385
Db 3393 TGGAGCTCAGGTGACCTGGATCTGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3452
QY 1386 CCCTGGCTACCTGGACAACTGGTCAAGATGGTAAGCTTGATATCAGGGAATTCAGG 1445
Db 3453 AGTTGGCTCACCAGGAGAACTGGGTGAGGAGTACAGCTGGCCCAATTTGGTTTACCAGG 3512
```

4156 GGAATCCCGAGAGAAAGGCAAGAGGCGCAGAGATGGAAAGCCGGGTCCCTCGAGAG 4215
 1747 AAGGAGAGAGAGAGATCCCTCGTGTCTCCGGGCGAGGATGGAAACACGGGAGAGCTCGA 1806
 4216 CCGGCAAGAGAGAGAGAGGCTTACAGAGCCAGAGGGTGCAGGAGCCCACTGGC 4275
 1807 ATCCAGAGATTTCTGGAAACCGAGGATTAATGGGCCAAAGGAGAAATTCGGCTCCA 1866
 4276 TTCAAGGACACACAGGCGATTCGTGTGACCCGGTCCCGGGAGAGTCTGGTGCATG 4335
 1867 GCACAGAGAGAGAGAGAGCCCGAGGATCCCTGGTTTAAATGGAGCAATGGCTCA 1926
 4336 GGGCTTCTCGTCAAGAGGGTTACAGAGAAAGATGGTGACACTGGACCTGGCCCA 4395
 1927 CCAGGCGAGCTGGAAACACCGGATCTAAGAGGAAGCAAAAGGTCAACTGGAATTCAGGG 1986
 4396 CAGGTCCTCCAGGACCAAGGGGCCACCGGCAAGAATGGATCACCGGATCTCCAGA 4455
 1987 ATGCTGGGGTCTTCTGGCTCAAGGAGAACACAGGAGCAACGGGTTCCCGAGAGACCA 2046
 4456 GAGCTTGGCCCTTCAGGAACCCCTGGCCAGAAAGCAAAAGGGGAAATGGCAGCCCA 4515
 2047 GATATCATGGTTTACCGGGATTCAGGAAAGAGGGGCAAGAGGAATCAAGTGAA 2106
 4516 GAGCTTCTGGCTTCTGGTCTCCCTGGGCTCCCGGAGAACAGGAGAGAGAGGTC 4575
 2107 AAGGTATTTCAGGGTCAAAAGGAGAGAAATGGAGACAGGGAATTCAGGGCAACAGGGA 2166
 4576 CAGGCAAGAGGGGGTCTCTGGAGACCTGGAGAGCTGGATTCAGAGAGAGAGGGA 4635
 2167 ATTCAGAGGCATCATGGTGAAGAGAGAGAGAGTGAAGAGGAGAACTGTGTCCGA 2226
 4636 GATCTGGGATCAAGAGTGAAGAGACCTCTCTGGTGAAGAGCCAGCTGGGACCT 4695
 2227 GGTGCCATTGGATCAAGAGAGATCTGGGTGGATGGCTTGATGGGGCCCGCAGGTCT 2286
 4696 GGAATCCAGGCGCACAAAGGCGCACAGGCTCTATGGGTCCCAAGGACTTACTGGGAG 4755
 2287 AAGGGCAACCTGGGATCCAGTCTCTCAGGACCCCGCAGGTTTGGATGGAGCCCGGA 2346
 4756 AATGGACAGTTGGACCCCGAGGCTCCAGGCGCAGCGGATTCAGGACTGAGGGG 4815
 2347 AGAGATTTTCAGAACAATTTATTCAGAAAGTTTGCACAGATGAATAGAGCCAGCTA 2406
 4816 GAGTCTCCATCCATGGAACCCCTCGTGGCTT-----ATTCAAGAGAGAGCTGGGAA 4868
 2407 CCAGTCTTACTTCAGAGTGAAGAAATTAGAAATTGTGATCTTGGCTGTCCCAACATGC 2466
 4869 GCAGCTTGAAACAGACTCGCTACTCTGGGCGAGATGCCCGCGGTACATGAAGTC 4928
 2467 -TCCCGGGTATTCTTGGGCGACCTGTGTCGATAGGCGCCAGAGGGTCCAGAGGATTACC 2525
 4929 ATCTAAGGAGACCTTGGGCGCCCGAGGCGCCCTGGAAAGATGGCTTCAGGCGGGC 4988
 2526 TGGTTTCCAGAGAGAGATGGTGTCTGTAGTGTGGTGTCTCTGGAGCTCCAGGTG 2585
 4989 CGGCCCCATGGGGGACAGGTCTGTGGGAGGGGGTCTGGAAGGACCCCTCTGGAC 5048
 2586 CAGAGGATTAAGAGGCTTACAGAGAAATGGGAAAGAGGAGCCAGGGTTTCGGTA 2645
 5049 CATAGTCCCAAGGTGAGGAGGAGCAAGGTGACCCAGGTGACCTGGAGTTGGCT 5108
 2646 TCCTGGAGAAAGAGTCTCTGTGTCCTCCAGGTTCAGAGGGCCCTCTGGAATAAGCAA 2705
 5109 CCGAGGAGAGTGGACCCCTGGAATCCAGGTCAACCCGGGAGCTTGGCTATGCTAA 5168
 2706 AGAAGTCTCCAGAGAGCCAGGTCTCCCTGGCAAGATGGAGACCATGGAAAACTGG 2765
 5169 AGATGGACTTCTGGGATCTGGGCTCCAGGGGAGACAGGACAGCTGGACATCTGG 5228
 2766 AATCCAGGCGCAACCGGCGCCCGCAGGCATCTGCGACCCCATCACTGT 2814
 5229 CTTCCAGGACCTCCCGTCCCCCAGGCCAATGTGACCCCTTCCAGTGT 5277

RESULT 6
 US-09-918-995-17490
 ; Sequence 17490, Application US/09918995
 ; Publication No. US20030073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; FROM VARIOUS CDNA LIBRARIES
 ; FILE REFERENCE: 20411-756
 ; CURRENT APPLICATION NUMBER: US/09/918,995
 ; CURRENT FILING DATE: 2001-07-30
 ; PRIOR APPLICATION NUMBER: US/09/235,076
 ; PRIOR FILING DATE: 1999-01-20
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 17490
 ; LENGTH: 496
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)...(496)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-918-995-17490
 Query Match 9.3%; Score 267; DB 11; Length 496;
 Best Local Similarity 100.0%; Pred. No. 2.3e-63;
 Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2599 GGCCTACAGAGAAATGGGAAAGAGGAGCCCAAGGTTTGGGTATCTCGAGAACAA 2658
 DB 85 GGCCTACAGAGAAATGGGAAAGAGGAGCCCAAGGTTTGGGTATCTCGAGAACAA 144
 QY 2659 GGTCTCTCTGTCCTCCAGAGTCCAGAGGGCCCTCTGGAATAAGCAAGAGTCTCTCA 2718
 DB 145 GGTCTCTCTGTCCTCCAGAGTCCAGAGGGCCCTCTGGAATAAGCAAGAGTCTCTCA 204
 QY 2719 GGAGACCCAGGTCTCTCTGGCAAGAGTGGAGACCATGGAACCTGGAATCCAGGGCAA 2779
 DB 205 GGAGACCCAGGTCTCTCTGGCAAGAGTGGAGACCATGGAACCTGGAATCCAGGGCAA 264
 QY 2779 CGAGCCCTCCAGGATCTCGACCCATCACTATGTTTGTAGTAAATTCGAGAGAGAT 2838
 DB 265 CGAGCCCTCCAGGATCTCGACCCATCACTATGTTTGTAGTAAATTCGAGAGAGAT 324
 QY 2839 CCGTTTCAGAAAGGACCAACTATTAG 2865
 DB 325 CCGTTTCAGAAAGGACCAACTATTAG 351
 RESULT 7
 US-10-198-846-13890
 ; Sequence 13890, Application US/10198846
 ; Publication No. US2003009974A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liilie, James
 ; APPLICANT: Xu, Yongyao
 ; APPLICANT: Wang, Youzhen
 ; APPLICANT: Steinmann, Kathleen
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
 ; FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
 ; THERAPY OF BREAST CANCER
 ; FILE REFERENCE: MRI-049
 ; CURRENT APPLICATION NUMBER: US/10/198,846
 ; CURRENT FILING DATE: 2002-07-18
 ; PRIOR APPLICATION NUMBER: 60/306,220
 ; PRIOR FILING DATE: 2001-07-18
 ; NUMBER OF SEQ ID NOS: 14094
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13890
 ; LENGTH: 6545


```
QY 2224 CGAGTGCCTTGGATCAAAAGGAGAAATCTGGGTGGATGGCTTGTATGGGGCCCGCAGGT 2283
Db 511 CGAGTGCCTTGGATCAAAAGGAGAAATCTGGGTGGATGGCTTGTATGGGGCCCGCAGGT 570
QY 2284 CTAAGGGGCAACCTGGGGATCCAGGTCTCAGGGACCCCGAGGTTTGGATGGGAAGCCC 2343
Db 571 CTAAGGGGCAACCTGGGGATCCAGGTCTCAGGGACCCCGAGGTTTGGATGGGAAGCCC 630
QY 2344 GGAAGAGAGTTTTCAGAACAAATTTATTCCGACAAATTTTGCACAGATGTAATAAGAGCCAG 2403
Db 631 GGAAGAGAGTTTTCAGAACAAATTTATTCCGACAAATTTTGCACAGATGTAATAAGAGCCAG 690
QY 2404 CTACAGCTTTACTTCAGAGTGGAGAAATTAAGAAATTTGATCATTCCTGTCCCAACAT 2463
Db 691 CTACAGCTTTACTTCAGAGTGGAGAAATTAAGAAATTTGATCATTCCTGTCCCAACAT 750
QY 2464 GGCTCCCGGGTATTCCTGGGCCACCTGGTCGGATAGCCAGAGGGTCCAGAGGATTA 2523
Db 751 GGCTCCCGGGTATTCCTGGGCCACCTGGTCGGATAGCCAGAGGGTCCAGAGGATTA 810
QY 2524 CCTGGTTTCCAGGAAGAGATGGTGTCTTCCGATTTAGTGGGTGTCCTCGACGTCCAGGT 2583
Db 811 CCTGGTTTCCAGGAAGAGATGGTGTCTTCCGATTTAGTGGGTGTCCTCGACGTCCAGGT 870
QY 2584 GTCAGAGATTAAGAGGCTTACCAGGAAGAAATGGGAAAGAGGAGCCAAAGGTTTGGG 2643
Db 871 GTCAGAGATTAAGAGGCTTACCAGGAAGAAATGGGAAAGAGGAGCCAAAGGTTTGGG 930
QY 2644 TATCTCGAGAACAGGTCTCTCGTCCCGCCAGTCCAGAGGGCCCTCCTGGGAATAGC 2703
Db 931 TATCTCGAGAACAGGTCTCTCGTCCCGCCAGTCCAGAGGGCCCTCCTGGGAATAGC 990
QY 2704 AAAGAAGTCTCTCCAGGAGACCCAGGTCTCCTGGCAAGATGAGACCAATGGAAACCT 2763
Db 991 AAAGAAGTCTCTCCAGGAGACCCAGGTCTCCTGGCAAGATGAGACCAATGGAAACCT 1050
QY 2764 GGAATCCAGGGCAACAGGCCCCCAGGATCTGCGACCATCACTATGTTTAGTGTA 2823
Db 1051 GGAATCCAGGGCAACAGGCCCCCAGGATCTGCGACCATCACTATGTTTAGTGTA 1110
QY 2824 ATTGCCAAGAGATCCCTTCAGAAAGGACCAAACTATTAG 2865
Db 1111 ATTGCCAAGAGATCCCTTCAGAAAGGACCAAACTATTAG 1152
```

RESULT 3

```
US-09-918-995-33525
; Sequence 33525, Application US/09918995
; Publication No. US20030073623A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33525
; LENGTH: 412
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(412)
; OTHER INFORMATION: n = A,T,C or G
```

US-09-918-995-33525

Query Match 13.3%; Score 381; DB 11; Length 412;

Best Local Similarity 97.8%; Pred. No. 4.3e-95;

```
Matches 400; Conservative 0; Mismatches 0; Indels 9; Gaps 1;
QY 1103 AAAAAATTGAAAAAAGCCCTTACATCCAGTTTAGGGATCTTGATCAATGGGAAACCC 1162
Db 3 AAAAAATTGAAAAAAGCCCTTACATCCAGTTTAGGGATCTTGATCAATGGGAAACCC 62
QY 1163 AAATTGGAAATATTCTGGAAAGAAAGAACTGTTCAGTTTGATGTCGAAAGTTGCGAA 1222
Db 63 AAATTGGAAATATTCTGGAAAGAAAGAACTGTTCAGTTTGATGTCGAAAGTTGCGAA 122
QY 1223 TCTACTGTGACCCCAAGACAGAACCGGAGACAGCATGTGAGATTCCTCGATT 1278
Db 123 TCTACTGTGACCCCAAGACAGAACCGGAGACAGCATGTGAGATTCCTCGATTAAATG 182
QY 1279 -----TGCCTTAATGTCCTCCAGTGTAGGTTCAACTCCAGCTCCCTGTGATTTGTCCTC 1333
Db 183 GAGAGTGCCTTAATGTCCTCCAGTGTAGGTTCAACTCCAGCTCCCTGTGATTTGTCCTC 242
QY 1334 CGGGAAGACAGGACTTCAAGSCCCCAAGGTGACCTCGGACTGCCTGGGAACCCCTGCT 1393
Db 243 CGGGAAGACAGGACTTCAAGSCCCCAAGGTGACCTCGGACTGCCTGGGAACCCCTGCT 302
QY 1394 ACCCTGGACAACTCTGTCAGATGTTAGCCCTGGATATCAGGAAATTCAGGGACACCAG 1453
Db 303 ACCCTGGACAACTCTGTCAGATGTTAGCCCTGGATATCAGGAAATTCAGGGACACCAG 362
QY 1454 GTGTTCCAGGATCTCCAGGAATACAGGAGCTCGAGGACTACCAAGTTA 1502
Db 363 GTGTTCCAGGATCTCCAGGAATACAGGAGCTCGAGGACTACCAAGTTA 411
```

RESULT 4

US-10-202-167-3

```
; Sequence 3, Application US/10202167
; Publication No. US20030143564A1
```

; GENERAL INFORMATION:

```
; APPLICANT: Burgeson, Robert Eugene
; APPLICANT: Koch, Manuel
; APPLICANT: Bruckner-Tuderman, Leena
; APPLICANT: Keene, Douglas R.
; APPLICANT: Brunken, William Joseph
; TITLE OF INVENTION: COLLAGEN XXII, A NOVEL HUMAN COLLAGEN
; TITLE OF INVENTION: AND USES THEREOF
```

; FILE REFERENCE: 10287-072001

; CURRENT APPLICATION NUMBER: US/10/202,167

; PRIOR FILING DATE: 2003-03-31

; PRIOR APPLICATION NUMBER: US 60/309,158

; PRIOR FILING DATE: 2001-07-31

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 4881

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-202-167-3

Query Match 9.9%; Score 285; DB 12; Length 4881;

Best Local Similarity 50.4%; Pred. No. 1.3e-67;

Matches 751; Conservative 0; Mismatches 730; Indels 8; Gaps 2;

```
QY 1327 TGTCTCTCCGGGAAAAACCAAGGACTTCAAGGCCCCCAAGAGTGACCTCGGACTGCTGGGAAC 1386
Db 3349 TGCCTCTCTGGCCCCCAGGCTCTCCCTGCTTACCAGTTTAAAGGGGACAAAGGTGTC 3408
QY 1387 CTGGCTACCTTGGACCAACCTGTCAGATGTAAGCCCTGGATATCAGGAAATTCAGAGG 1446
Db 3409 CCAGGAAGCCAGGAGAGAGGACAGAAAGGAAAAAGGAGAGGCTGGGCTCCAGGC 3468
QY 1447 ACACGAGGTGTTCCAGGATCTCCAGGAATACAGGAGCTCGAGGACTACAGGTTACAAA 1506
Db 3469 CTACGAGGSCCCCGAGGAATAGCTGGACCAAGGGAAGTCAAGGAGAACTGGTGCAGAT 3528
QY 1506 GAGAGACCAAGGCGAGATGTTGACAAAGGTGATCTGGACTTCTCTGGTTTCTCTGGGCTT 1566
```

```
QY 1824 AAACCGAGGATTAATGGGCCCAAAAGGAGAAATTTGGGCTCCAGGACAGCAAGGAAAAA 1883
DB |||||||
QY 249 AAAACGAGGATTAATGGGCCAAAGGAGAAATTTGGGCTCCAGGACAGCAAGGAAAAA 308
DB |||||||
QY 1884 AGGAGCCCCAGGATGCTGTTTAAATGGGAAGCAATGGCTACAGGCGACGCTCGAAC 1943
DB |||||||
QY 309 AGGAGCCCCAGGATGCTGTTTAAATGGGAAGCAATGGCTACAGGCGACGCTCGAAC 368
DB |||||||
QY 1944 ACCGGATCTAAGGGAACCAAGGTGAACCTTGAATTCAGGGATGCTGGGGCTCTGG 2003
DB |||||||
QY 369 ACCGGATCTAAGGGAACCAAGGTGAACCTTGAATTCAGGGATGCTGGGGCTCTGG 428
DB |||||||
QY 2004 GCTCAAGGGAACCAAGGATGCTTCCCGAGGAAACCAAGGATGCTGGGGCTTACC 2063
DB |||||||
QY 429 GCTCAAGGGAACCAAGGATGCTTCCCGAGGAAACCAAGGATGCTGGGGCTTACC 488
DB |||||||
QY 2064 CGGGATTCAGGGAACCAAGGATGCTTCCCGAGGAAACCAAGGATGCTGGGGCTTACC 2123
DB |||||||
QY 489 CGGGATTCAGGGAACCAAGGATGCTTCCCGAGGAAACCAAGGATGCTGGGGCTTACC 548
DB |||||||
QY 2124 AAAGGAGAAATGGAACAGAGGAAATTCAGGGCAACAGGAAATTCAGGGCAATCGG 2183
DB |||||||
QY 549 AAAGGAGAAATGGAACAGAGGAAATTCAGGGCAACAGGAAATTCAGGGCAATCGG 608
DB |||||||
QY 2184 TGCAAAAGGAGAGAGGTGAAAGGGAGAACCTGGTCTCCGAGGTGCAATGGATCAAA 2243
DB |||||||
QY 609 TGCAAAAGGAGAGAGGTGAAAGGGAGAACCTGGTCTCCGAGGTGCAATGGATCAAA 668
DB |||||||
QY 2244 AGGAGAAATCTGGGTGATGCTTGAATGGGCGCGAGGCTTAAAGGGCAACCTGGGA 2303
DB |||||||
QY 669 AGGAGAAATCTGGGTGATGCTTGAATGGGCGCGAGGCTTAAAGGGCAACCTGGGA 728
DB |||||||
QY 2304 TCCAGGTCTCAGGACCCCGAGGTTGGATGGGAAGCCCGAGAGAGAGTTTTCAGAAC 2363
DB |||||||
QY 729 TCCAGGTCTCAGGACCCCGAGGTTGGATGGGAAGCCCGAGAGAGAGTTTTCAGAAC 788
DB |||||||
QY 2364 ATTTATTCAGAGTTTGCACAGATGTAATAGAGCCCGAGCTACAGCTTACTTTCAGAG 2423
DB |||||||
QY 789 ATTTATTCAGAGTTTGCACAGATGTAATAGAGCCCGAGCTACAGCTTACTTTCAGAG 848
DB |||||||
QY 2424 TGGAGAAATAGAAATTTGATGATCTGCTGCCAATGCTGCTCCCGGCTTCTCTGG 2483
DB |||||||
QY 849 TGGAGAAATAGAAATTTGATGATCTGCTGCCAATGCTGCTCCCGGCTTCTCTGG 908
DB |||||||
QY 2484 GCCACCTGTCGATAGCCCGAGAGGTTCCAGAGSATTACCTGGTTTCCAGGAAGAGA 2543
DB |||||||
QY 909 GCCACCTGTCGATAGCCCGAGAGGTTCCAGAGSATTACCTGGTTTCCAGGAAGAGA 968
DB |||||||
QY 2544 TGGTGTCTCGATAGTGGGTGCTCCCTGGAGTCCAGGTGTCAGAGATTTAAAGGCT 2623
DB |||||||
QY 969 TGGTGTCTCGATAGTGGGTGCTCCCTGGAGTCCAGGTGTCAGAGATTTAAAGGCT 1028
DB |||||||
QY 2604 ACCAGGAAGAAATGGGGAAGGAGGCAAGGGTTTGGGTATCTCGAGAACCAAGGTCC 2663
DB |||||||
QY 1029 ACCAGGAAGAAATGGGGAAGGAGGCAAGGGTTTGGGTATCTCGAGAACCAAGGTCC 1088
DB |||||||
QY 2664 TCCTGTCTCCCGAGTCCAGAGGCTTCTCGAATAGCAAGAGTCTCTCAGGAGA 2723
DB |||||||
QY 1089 TCCTGTCTCCCGAGTCCAGAGGCTTCTCGAATAGCAAGAGTCTCTCAGGAGA 1148
DB |||||||
QY 2724 CCCAGGTCTCCCTGCAAGATGGAGACCATGGAAACCTGGAAATCCAGGGCAACAGG 2783
DB |||||||
QY 1149 CCCAGGTCTCCCTGCAAGATGGAGACCATGGAAACCTGGAAATCCAGGGCAACAGG 1208
DB |||||||
QY 2784 CCCCCAGGATCTCGGACCCATCACTATGTTTATGTTATTTGCGAAGAGATTCGGT 2843
DB |||||||
QY 1209 CCCCCAGGATCTCGGACCCATCACTATGTTTATGTTATTTGCGAAGAGATTCGGT 1268
DB |||||||
QY 2844 CAGAAAGGACCAACTATTAG 2865
DB |||||||
QY 1269 CAGAAAGGACCAACTATTAG 1290
DB |||||||
```

```
RESULT 2
US-10-037-270-282
; Sequence 282, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, fuding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No US20030104529A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pc FL_genes Version 1.0
; SEQ ID NO 282
; LENGTH: 2230
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCAT:CN: (85)...(1152)
US-10-037-270-282

Query Match 37.1%; Score 1062; DB 14; Length 2230;
Best Local Similarity 100.0%; Pred. No. 9.3e-284;
Matches 1062; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1804 GGAATCCCAAGGATTTCTCTGAAACCCAGGATTAATGGGCCAAAGGAGAAATTTGGGCT 1863
DB |||||||
QY 91 GGAATCCCAAGGATTTCTCTGAAACCCAGGATTAATGGGCCAAAGGAGAAATTTGGGCT 150
DB |||||||
QY 1864 CCAGGACAGCAAGGAAAAAAGAGGCCCCAGGATGCTGCTGGTTTAAATGGGAAGCAATGGC 1923
DB |||||||
QY 151 CCAGGACAGCAAGGAAAAAAGAGGCCCCAGGATGCTGCTGGTTTAAATGGGAAGCAATGGC 210
DB |||||||
QY 1924 TCACCAAGGCGCTCGAACACCGGATCTAAGGGAAGCAAGGATTAATGGGAATTCAA 1983
DB |||||||
QY 211 TCACCAAGGCGCTCGAACACCGGATCTAAGGGAAGCAAGGATTAATGGGAATTCAA 270
DB |||||||
QY 1984 GGGATCCCTGGGCTTCTGGGCTCAAGGGAGAACCCAGGAGCAACCGGTTCCCGAGGAA 2043
DB |||||||
QY 271 GGGATCCCTGGGCTTCTGGGCTCAAGGGAGAACCCAGGAGCAACCGGTTCCCGAGGAA 330
DB |||||||
QY 2044 CCAGGATACATGGTTTACCCGGGATTCAGGAAAAAAGGGGGCAAGGAAATCAAGGT 2103
DB |||||||
QY 331 CCAGGATACATGGTTTACCCGGGATTCAGGAAAAAAGGGGGCAAGGAAATCAAGGT 390
DB |||||||
QY 2104 GAAAAAGGTATTTCAGGCTCAAAAGGAGAAATTCAGGAAACAGGAAATTCAGGGAACAG 2163
DB |||||||
QY 391 GAAAAAGGTATTTCAGGCTCAAAAGGAGAAATTCAGGAAACAGGAAATTCAGGGAACAG 450
DB |||||||
QY 2164 GGAATTCAGGGCCATCATGGTGC AAAAGGAGAGAGAGGTGAAAAGGAGAACCTGGTGTCT 2223
DB |||||||
QY 451 GGAATTCAGGGCCATCATGGTGC AAAAGGAGAGAGAGGTGAAAAGGAGAACCTGGTGTCT 510
DB |||||||
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 19:13:33 ; Search time 702 Seconds
(without alignments)
10410.575 Million cell updates/sec

Title: US-09-996-611B-5

Perfect score: 2865

Sequence: 1 atgctcactattacatt.....gaaagaccacacatttag 2865

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 1708419 seqs, 1275431651 residues

Total number of hits satisfying chosen parameters: 3416838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	1282	44.7	2375	10	US-09-764-864-232
2	1062	37.1	2230	14	US-10-037-270-282
3	381	13.3	412	11	US-09-918-995-33525
4	285	9.9	4881	12	US-10-202-167-3
5	285	9.9	6352	12	US-10-202-167-1
6	267	9.3	496	11	US-09-918-995-17490
7	244.6	8.5	6545	14	US-10-198-846-13890
8	243	8.5	6281	14	US-10-084-817-155
9	242.6	8.5	5460	12	US-10-301-822-32
10	242.6	8.5	5460	14	US-10-177-293-67
11	242.6	8.5	5489	12	US-10-096-534-13
12	242.6	8.5	8664	14	US-10-198-846-9949
13	241.4	8.4	6319	12	US-10-096-534-15
14	239.8	8.4	6158	10	US-09-919-497-6
15	239.8	8.4	6158	10	US-09-954-456-762
16	239.8	8.4	6158	12	US-09-873-367C-432

17	233	8.1	4629	12	US-10-293-582-22	Sequence 22, Appl
18	233	8.1	4629	14	US-10-177-293-69	Sequence 69, Appl
19	233	8.1	6217	12	US-10-301-822-34	Sequence 34, Appl
20	231.8	8.1	4664	12	US-10-194-441A-2	Sequence 2, Appl
21	231.5	8.1	5060	14	US-10-175-523-181	Sequence 181, Appl
22	227.2	7.9	4660	12	US-10-133-013-132	Sequence 132, Appl
23	227.2	7.9	4660	14	US-10-084-817-154	Sequence 154, Appl
24	227.2	7.9	4837	14	US-10-084-817-45	Sequence 45, Appl
25	222.8	7.8	8664	14	US-10-198-846-9949	Sequence 9949, Ap
26	222.6	7.8	3671	14	US-10-084-817-67	Sequence 67, Appl
27	219.6	7.7	2511	14	US-10-157-031-61	Sequence 61, Appl
28	215	7.5	2542	10	US-09-964-824A-255	Sequence 255, App
29	215	7.5	2542	10	US-09-954-531-961	Sequence 961, App
30	212.8	7.4	6200	12	US-09-795-061-3	Sequence 3, Appl
31	211.4	7.4	6109	12	US-09-795-061-1	Sequence 1, Appl
32	207.8	7.3	7924	14	US-10-198-846-9752	Sequence 9752, Ap
33	207.6	7.2	5274	14	US-10-198-846-13840	Sequence 13840, A
34	206	7.2	5086	10	US-09-880-107-3947	Sequence 3947, Ap
35	206	7.2	5086	12	US-09-873-319-745	Sequence 745, App
36	206	7.2	5086	12	US-09-960-706-1123	Sequence 1123, Ap
37	206	7.2	5086	12	US-10-101-510-24	Sequence 24, Appl
38	206	7.2	5086	12	US-10-301-822-30	Sequence 30, Appl
39	206	7.2	5086	12	US-09-873-367C-522	Sequence 522, App
40	206	7.2	5086	12	US-09-873-367C-1067	Sequence 1067, Ap
41	206	7.2	5086	14	US-10-171-311-37	Sequence 37, Appl
42	206	7.2	5451	12	US-10-101-510-597	Sequence 597, App
43	205.8	7.2	4812	14	US-10-084-817-4	Sequence 4, Appl
44	205.8	7.2	5416	10	US-09-954-456-786	Sequence 786, App
45	205.8	7.2	5416	10	US-09-880-107-2094	Sequence 2094, Ap

ALIGNMENTS

RESULT 1

US-09-764-864-232
; Sequence 232, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT Filing DATE: 2001-01-17
; Prior application data removed - consult: PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 232
; LENGTH: 2375
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-864-232

Query Match 44.7%; Score 1282; DB 10; Length 2375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	1584	AAAGGTGAAATGGGTGGCCAAAGGAGACAAAGGATCACCTGGATTATGCGAAAAGGG	1643
Db	9	AAAGGTGAAATGGGTGGCCAAAGGAGACAAAGGATCACCTGGATTATGCGAAAAGGG	68
Oy	1640	TGCAAAAGGTGAAAGGGGATGCTGGCTTCCTGGCTCCCTGGACCTGCTGGAGAAC	1703
Db	69	TGCAAAAGGTGAAAGGGGATGCTGGCTTCCTGGCTCCCTGGACCTGCTGGAGAAC	128
Oy	1704	AGGAACATGAAAGGATGGATTAATGGGTAGTCCCGTTCAGGGAGAGAGAGGATC	1763
Db	129	AGGAACATGAAAGGATGGATTAATGGGTAGTCCCGTTCAGGGAGAGAGAGGATC	188
Oy	1764	CCCTGTGCTCCGGSCAGGATGGAAACACGGGAGAGCTCGGAATCCAGGATTCCTG	1823
Db	189	CCCTGTGCTCCGGSCAGGATGGAAACACGGGAGAGCTCGGAATCCAGGATTCCTG	249

Job time : 10163 secs

c r

c r

* 62518 71007: contig of 8490 bp in length
 * 71008 71107: gap of 100 bp
 * 71108 79805: contig of 8698 bp in length
 * 79805 79905: gap of 100 bp
 * 79905 81160: contig of 1255 bp in length
 * 81160 81260: gap of 100 bp
 * 81260 92353: contig of 11093 bp in length
 * 92353 92453: gap of 100 bp
 * 92453 101125: contig of 8672 bp in length
 * 101125 101226: gap of 100 bp
 * 101226 117073: contig of 15848 bp in length
 * 117073 117174: gap of 100 bp
 * 117174 136020: contig of 18847 bp in length
 * 136020 136121: gap of 100 bp
 * 136121 161828: contig of 25708 bp in length
 * 161828 161929: gap of 100 bp
 * 161929 187999: contig of 26071 bp in length.

FEATURES

source
 1. .187999
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1"
 /clone="RP11-18A3"
 /clone_lib="RPC1-11 Human Male BAC"
 1. .1227
 /note="assembly_fragment"
 1328. .2416
 /note="assembly_fragment"
 2517. .4057
 /note="assembly_fragment"
 4158. .5171
 /note="assembly_fragment"
 5272. .6380
 /note="assembly_fragment"
 6481. .7659
 /note="assembly_fragment"
 7760. .9214
 /note="assembly_fragment"
 9315. .10932
 /note="assembly_fragment"
 11033. .12035
 /note="assembly_fragment"
 12136. .13215
 /note="assembly_fragment"
 13316. .14645
 /note="assembly_fragment"

Query Match 0.8%; Score 22; DB 2; Length 187999;
 Best Local Similarity 100.0%; Pred. No. 7.6;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1175 ATTCTGGAAGAGAACTGT 1196
 Db 64451 ATTCTGGAAGAGAACTGT 64430

RESULT 45

CNS05TDR 188162 bp DNA linear PRI 26-APR-2001
 LOCUS Human chromosome 14 DNA sequence BAC R-1079H9 of library RPC1-11
 DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.

ACCESSION AL357092
 VERSION AL357092.4 GI:12711298
 KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 188162)
 AUTHORS Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P., Brottier, P., Cattolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,

Levy, M., Eckerberg, R., Bruls, T., deBerardinis, V., Gruaud, C., Gyapay, G., Saurin, W. and Weissenbach, J.
 Sequencing of the human chromosome 14
 Unpublished
 2 (bases 1 to 188162)
 Genoscope.
 Direct Submission
 Submitted (26-APR-2001) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr
 On Feb 7, 2001 this sequence version replaced gi:9756028.
 ----- Genome Center
 Center: Genoscope / Centre National de Sequencage
 Center code: GS
 Web site: http://www.genoscope.cns.fr/
 Contact: SeqRef@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
 Upstream BAC (overlapping the T7 end) : C-2341M24 (AC=AL445593)
 Downstream BAC (overlapping the SP6 end) : R-369N15 -----
 Summary Statistics
 Assembly program: Phrap; version 2.0
 Quality coverage: 8.14x in Q20 bases; sum-of-contigs

Overall quality chart :
 Range : bases
 C :
 1 - 9 :
 10 - 19 : 6
 20 - 29 : 13
 30 - 39 : 48
 40 - 49 : 121
 50 - 59 : 4130
 60 - 69 : 5684
 70 - 79 : 17030
 80 - 89 : 53373
 90 - 99 : 106067

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES

source
 1. .188162
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone="R-1079H9"
 /clone_lib="RPC1-11"
 24911. .25016
 /note="matching EMBL:G41935
 RHdb:RH100302
 RHdb:RH94880
 dbSTS:STS67871
 Identified using the e-PCR software (G. Schuler)"
 164360. .164606
 /note="matching EMBL:Z51730
 RHdb:RH97872
 RHdb:RH31322
 dbSTS:STS6005
 Identified using the e-PCR software (G. Schuler)"
 BASE COUNT 61922 a 34314 c 34214 g 57712 t
 ORIGIN

Query Match 0.8%; Score 22; DB 9; Length 188162;
 Best Local Similarity 100.0%; Pred. No. 7.6;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2123 AAAAGGGGAGAAATGGAAGACA 2144

Db 135221 AAAAGGGGAGAAATGGAAGACA 135242

Search completed: October 3, 2003, 03:02:03


```

SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 187999)
REFERENCE
AUTHORS
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 1, clone RP11-18A3
UNPUBLISHED
2 (bases 1 to 187999)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Bouckgatter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Govette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamarques, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrim, J., Menues, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Navlor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 187999)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Bouckgatter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Govette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamarques, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrim, J., Menues, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Navlor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 22, 2000 this sequence version replaced gi:7328793.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/BM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7568
Center clone name: 18_A_3
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

```

```

Assembly program: Phrap; version 0.96c731
Consensus quality: 167725 bases at least Q40
Consensus quality: 178025 bases at least Q30
Consensus quality: 182047 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 184399; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
-----
1
1227: contig of 1227 bp in length
1228 1327: gap of 100 bp
1328 2416: contig of 1089 bp in length
2417 2516: gap of 100 bp
2517 4057: contig of 1541 bp in length
4058 4157: gap of 100 bp
4158 5171: contig of 1014 bp in length
5172 5271: gap of 100 bp
5272 6380: contig of 1109 bp in length
6381 7659: contig of 1179 bp in length
7660 7759: gap of 100 bp
7760 9214: contig of 1455 bp in length
9215 9315: gap of 100 bp
10313 10932: contig of 1618 bp in length
10933 12035: contig of 1003 bp in length
12036 12135: gap of 100 bp
12136 13215: contig of 1080 bp in length
13216 13315: gap of 100 bp
13316 14645: contig of 1330 bp in length
14646 14745: gap of 100 bp
14746 15925: contig of 1180 bp in length
15926 16025: gap of 100 bp
16026 17158: contig of 1133 bp in length
17159 17258: gap of 100 bp
17259 19445: contig of 2187 bp in length
19446 19545: gap of 100 bp
19546 21242: contig of 1697 bp in length
21243 21342: gap of 100 bp
21343 24035: contig of 2693 bp in length
24036 25142: contig of 1007 bp in length
25143 25242: gap of 100 bp
25243 27117: contig of 1875 bp in length
27118 27217: gap of 100 bp
27218 28912: contig of 1695 bp in length
28913 29012: gap of 100 bp
29013 31508: contig of 2496 bp in length
31509 31608: gap of 100 bp
31609 32697: contig of 1089 bp in length
32698 32797: gap of 100 bp
32798 32798: contig of 4478 bp in length
32799 37375: gap of 100 bp
37376 40982: contig of 3685 bp in length
40983 41080: gap of 100 bp
41081 45215: contig of 4135 bp in length
45216 45315: gap of 100 bp
45316 48710: contig of 3395 bp in length
48711 48810: gap of 100 bp
48811 52850: contig of 4040 bp in length
52851 52950: gap of 100 bp
52951 57072: contig of 4122 bp in length
57073 57172: gap of 100 bp
57173 62417: contig of 5245 bp in length
62418 62517: gap of 100 bp

```

```

misc_feature      82998..122831
/note="assembly_name:Contig11
clone_end:T7
vector_side:right"
misc_feature      122932..176694
/note="assembly_name:Contig12"
BASE COUNT      48989 a 36992 c 36355 g 53458 t      900 others
ORIGIN
Query Match      0.8%; Score 22; DB 2; Length 176694;
Best Local Similarity 100.0%; Pred.No. 7.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1862 CTCAGGACAGCAGGAGGAAAAA 1883
|||||
Db      135116 CTCAGGACAGCAGGAGGAAAAA 135137

RESULT 42
AL663099/c
LOCUS      AL663099      177135 bp      DNA      linear      ROD 19-JUL-2002
DEFINITION  Mouse DNA sequence from clone DN-384L23 on chromosome 3, complete
sequence.
ACCESSION  AL663099
VERSION    AL663099.7 GI:20068691
KEYWORDS   HTG.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 177135)
AUTHORS   Clark, G.
TITLE     Direct Submission
JOURNAL   Submitted (18-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
clonerequest@sanger.ac.uk
humquery@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:20067508.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
-----
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
DN-384L23 is from a DIL (Diabetes and Inflammation Laboratory) NOD
Mouse BAC library
VECTOR: pBACe3.6.
-----
Location/Qualifiers
1. .177135
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="3"
/clone="DN-384L23"
/clone_lib="NOD mouse library"
BASE COUNT      52554 a 40202 c 38941 g 45438 t

```

```

ORIGIN
Query Match      0.8%; Score 22; DB 10; Length 177135;
Best Local Similarity 100.0%; Pred.No. 7.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      23480 GAGAGTCTTCAGACAAATTAT 2369
|||||
Db      36082 GAGAGTCTTCAGACAAATTAT 36061

RESULT 43
AC122397
LOCUS      AC122397      181742 bp      DNA      linear      ROD 02-MAR-2003
DEFINITION  Mus musculus chromosome 15 clone RP24-90D21, complete sequence.
ACCESSION  AC122397
VERSION    AC122397.2 GI:28630148
KEYWORDS   HTG.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 181742)
AUTHORS   McPherson, J.D. and Waterston, R.H.
TITLE     The sequence of Mus musculus clone
Unpublished
REFERENCE  2 (bases 1 to 181742)
AUTHORS   McPherson, J.D. and Waterston, R.H.
TITLE     Direct Submission
JOURNAL   Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE  3 (bases 1 to 181742)
AUTHORS   McPherson, J.D. and Waterston, R.H.
TITLE     Direct Submission
JOURNAL   Submitted (02-MAR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT    On Mar 2, 2003 this sequence version replaced gi:21105847.
-----
Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
-----
Project Information
Center project name: M.BB0090D21
-----
Location/Qualifiers
1. .181742
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP24-30D21"
BASE COUNT      48193 a 43416 c 42696 g 47437 t
ORIGIN
Query Match      0.8%; Score 22; DB 10; Length 181742;
Best Local Similarity 100.0%; Pred.No. 7.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1507 GGAGAACGAGCGCGAGATGGTG 1528
|||||
Db      21276 GGAGAACGAGCGCGAGATGGTG 21297

RESULT 44
AC026928/c
LOCUS      AC026928      187999 bp      DNA      linear      HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 1 clone RP11-18A3 map 1, WORKING DRAFT
SEQUENCE, 37 unordered pieces.
ACCESSION  AC026928
VERSION    AC026928.2 GI:9369514
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H.NQ237E05
Summary Statistics
Sequencing vector: M13, 100%

```

sequencing vector: pTZ19, 1990
sequencing vector: plasmid:

```

Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Assembly: Dye-terminator Big Dye; 3% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 172839 bases at least Q43
Consensus quality: 173957 bases at least Q30
Consensus quality: 174626 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 175794; sum-of-ctnigs
Quality coverage: 5.63 in Q20 bases; agarose-fp
Quality coverage: 5.69 in Q20 bases; sum-of-ctnigs

* NOTE: This is a working draft sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the ctnigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

* 1 1659: contig of 1659 bp in length

* 1660 1759: gap of unknown length

* 1760 6808: contig of 5049 bp in length

* 6809 gap of unknown length

accrues

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-237E5"

```

```
feature
1. 1659
/note="assembly_name:Contig3"
feature
1760. 6808
/note="assembly_name:Contig4"
feature
6909. 13807
```

```

/note="assembly_name:Contig5"
13908. .20324
/note="assembly_name:Contig6"

```

feature 20425. 27826

```
/note="assembly_name:Contig7"
```

feature 27927. 36369

```
/note="assembly_name:Contiga
```

circle_end: 526
vector_circle_pos = "

```
feature      50727
vector_size:1ell 36470
```

```

2010.10.20 15:07:27
/note="assembly name:Contig9"

```

feature 50828. 82897

```
/note="assembly_name:Contig10"
```

```

misc_feature 497. 543
/notes="single clone coverage"
misc_feature 14367. 14393
/notes="single clone coverage"
misc_feature 47455. 47504
/notes="single clone coverage"
misc_feature 54353. 54525
/notes="single clone coverage"
misc_feature 55101. 55135
/notes="single clone coverage"
misc_feature 63551 63652
/notes="bacterial transposon excised; IS2 sequence can be
found in GenBank Accession Number AF000143.1 nucleotides
1305-2635; 1331 bp"
misc_feature 76511. 76572
/notes="single clone coverage"
misc_feature 98071. 98107
/notes="single clone coverage"
misc_feature 130697. 130699
/notes="single clone coverage"
misc_feature 151137. 151299
/notes="single clone coverage"
misc_feature 151338. 151471
/notes="single clone coverage"
misc_feature 163358. 163377
/notes="single clone coverage"
misc_feature 163419. 163469
/notes="single clone coverage"
BASE COUNT 47092 a 33553 c 35099 g 51007 t
ORIGIN
Query Match 0.8%; Score 22; DB 10; Length 166751;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1183 AAGAGAGAACTGTTTCAGTTG 1204
|||||
Db 6429 AAGAGAGAACTGTTTCAGTTG 6450

RESULT 40
AC102838/c
LOCUS
DEFINITION AC102838 169947 bp DNA linear HTG 23-MAR-2003
Mus musculus clone RP24-317M2, WORKING DRAFT SEQUENCE, 9 unordered
pieces.
ACCESSION AC102838
VERSION GI:29164563
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 169947)
Birren,B., Nusbaum,C. and Lander,E.
Unpublished
Mus musculus, clone RP24-317M2
2 (bases 1 to 169947)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Baetien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Cooke,P., DeAngelio,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehotsky,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,

```

```

Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,B., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 169947)
Birren,B., Nusbaum,C., Lander,E., Aboueleil,A., Allen,N.,
Anderson,S., Archachli,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,B., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hageciat,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,N.,
Meldrim,C., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,B., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 23 2003 this sequence version replaced gi:22381848.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIER
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 119940
Center clone name: 317 M.2
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 166683 bases at least Q40
Consensus quality: 167466 bases at least Q30
Consensus quality: 167808 bases at least Q20
Insert size: 173000; agarose-fp
Insert size: 169147; sum-of-contigs
Quality coverage: 11.7 in Q20 bases; agarose-fp
Quality coverage: 12.0 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 35676: contig of 35676 bp in length
* 35677 35776: gap of 100 bp
* 35777 36492: contig of 716 bp in length
* 36493 36592: gap of 100 bp
* 36593 37265: contig of 673 bp in length
* 37266 37365: gap of 100 bp
* 37366 38939: contig of 1574 bp in length
* 38940 39039: gap of 100 bp

```

TITLE
JOURNALREFERENCE
AUTHORSTITLE
JOURNAL

COMMENT

```

/misc_feature      /chromosome="12"
                   /clones="RP11-22B23"
                   1. .2004
                   /note="overlaps bases 21010. .23013 of clone AC137766"
repeat_region     2636. .2726
                   /rpt_family="L2"
repeat_region     3337. .3642
                   /rpt_family="MER2"
repeat_region     3927. .4028
                   /rpt_family="L2"
repeat_region     complement(4477. .4535)
                   /rpt_family="L2"
repeat_region     complement(5459. .5882)
                   /rpt_family="MLTID"
repeat_region     complement(5906. .6036)
                   /rpt_family="FLAM_A"
repeat_region     6077. .6906
                   /rpt_family="L1ME1"
repeat_region     6932. .7036
                   /rpt_family="L1MB3"
repeat_region     7047. .7736
                   /rpt_family="L1MB3"
repeat_region     7872. .7913
                   /rpt_family="AT_rich"
repeat_region     7939. .8120
                   /rpt_family="L1ME3"
repeat_region     8121. .8414
                   /rpt_family="AluY"
repeat_region     8415. .8541
                   /rpt_family="L1MB3"
repeat_region     8542. .8729
                   /rpt_family="AluJo"
repeat_region     8786. .9242
                   /rpt_family="L1ME3"
repeat_region     10126. .10620
                   /rpt_family="L1MB3"
repeat_region     complement(11274. .11881)
                   /rpt_family="MIR"
repeat_region     13186. .13271
                   /rpt_family="MIR"
repeat_region     complement(13276. .13406)
                   /rpt_family="MIR"
repeat_region     complement(14360. .14818)
                   /rpt_family="LTR56"
repeat_region     complement(16338. .16431)
                   /rpt_family="L2"
repeat_region     16785. .16978
                   /rpt_family="MER101"
repeat_region     16979. .17423
                   /rpt_family="LTR46"
repeat_region     17424. .17488
                   /rpt_family="MER101"
repeat_region     complement(17600. .17702)
                   /rpt_family="L2"
repeat_region     18161. .18264
                   /rpt_family="MIR"
repeat_region     18459. .18561
                   /rpt_family="MIR"
repeat_region     18614. .18894
                   /rpt_family="AluDb"
repeat_region     18895. .18931
                   /rpt_family="(TAAA)n"
repeat_region     18932. .18944
                   /rpt_family="AluDb"
repeat_region     complement(20133. .20222)

```

```

Query Match      0.8; Score 22; DB 9; Length 162572;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1178 CTGGAAGAGAGAACTGTTCA 1199
|||||

```

```

Db      140464 CTGGAAGAGAGAACTGTTCA 140485

RESULT 39
AC087066
LOCUS   Rattus norvegicus clone RP31-194D8 strain Brown Norway, complete
DEFINITION
ACCESSION AC087066      166751 bp      DNA      linear      ROD 01-JUN-2001
VERSION   AC087066.3   GI:14277289
KEYWORDS  HTG.
SOURCE    Rattus norvegicus (Norway rat)
ORGANISM  Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
REFERENCE 1 (bases 1 to 166751)
AUTHORS  Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
          Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S.,
          Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E., Lee-Lin,S.-Q.,
          Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B., Masiello,C.,
          Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
          Shevchenko,I., Snyder,B., Stantropop,S., Thomas,J.W., Thomas,P.J.,
          Tjongson,E.B., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
          Wetherby,K.D., Zhang,L.-H. and Green,E.D.
          NISC Comparative Sequencing Initiative
          Unpublished
          2 (bases 1 to 166751)
          Green,E.D.
          Direct Submission
          Submitted (05-DEC-2000) NIH Intramural Sequencing Center, 8717
          Grovermont Circle, Gaithersburg, MD 20877, USA
          3 (bases 1 to 166751)
          Green,E.D.
          Direct Submission
          Submitted (17-MAY-2001) NIH Intramural Sequencing Center, 8717
          Grovermont Circle, Gaithersburg, MD 20877, USA
          4 (bases 1 to 166751)
          Green,E.D.
          Direct Submission
          Submitted (01-JUN-2001) NIH Intramural Sequencing Center, 8717
          Grovermont Circle, Gaithersburg, MD 20877, USA
          On Jun 1, 2001 this sequence version replaced gi:14140275.
          ----- Genome Center
          Center: NIH Intramural Sequencing Center
          Center code: NISC
          Web site: http://www.nisc.nih.gov
          Contact: nisc.mouses@hgri.nih.gov
          ----- Project Information
          Center project name: ahh
          Center clone name: 194D08

This sequence was finished as follows unless otherwise noted:
all regions were double-stranded, sequenced with an
alternate chemistry, or covered by high quality data
(i.e., paired quality >= 30); an attempt was made to resolve
all sequencing problems, such as compressions and repeats;
all regions were covered by at least one plasmid subclone
or more than one M13 subclone; and the assembly was confirmed
by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of
this clone unless otherwise noted. If there are overlapping
clones, the overlaps are noted in the beginning and end of
the Features section.
          Location/Qualifiers
          1. 166751
             /organism="Rattus norvegicus"
             /mol_type="genomic DNA"
             /strain="Brown Norway"
             /db_xref="taxon:10116"
             /clone="RP31-194D8"
             /clone_lib="RP31"
FEATURES
          source

```

Query Match	0.8%;	Score 22;	DB 2;	Length 157793;
Best Local Similarity	100.0%;	Pred. No. 7.7;		
Matches 22;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	2348	GAGAGTTTTCAGACAAATTTAT	2369
DB	125386	GAGAGTTTTCAGACAAATTTAT	125407

RESULT 38	
AC009533	
LOCUS	162572 bp DNA linear PRI 08-JAN-2003
DEFINITION	Homo sapiens 12 BAC RP11-22823 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION	AC009533
VERSION	AC009533.10
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbacia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carrroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emeling,S., Escott,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunatratne,P., Hale,S., Hamilton,K., Han,J., Harris,K., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,J., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S., Ogih,M., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shoshchari,N., Sissor,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabot,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R., Weinstock,G. and Gibbs,R.
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (27-AUG-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
FEATURES	Source

REFERENCE	1 (bases 1 to 162572)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (31-DEC-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	2 (bases 1 to 162572)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (08-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Dec 31, 2002 this sequence version replaced gi:6649266. INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc:help@bcm.tmc.edu

CLONE LENGTH:	This sequence does not necessarily represent the entire length of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.
ANNOTATION OF FEATURES:	STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts. Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences. Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.
SEQUENCING READ COVERAGE:	Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.
QUALITY OF INDIVIDUAL BASES:	This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.htm.
FEATURES	Location/Qualifiers

1..162572	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"

RESULT 37

AC102109 157793 bp DNA linear HTG 12-JUN-2003
 LOCUS Mus musculus clone RP23-113117, WORKING DRAFT SEQUENCE, 9 unordered
 DEFINITION pieces.
 AC102109 GI:31621376
 VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 157793)
 Mus musculus, clone RP23-113117

2 (bases 1 to 157793)
 Unpublished

REFERENCE

AUTHORS

Birren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Gande,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulle,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., Lapocque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
 Menueu,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
 Seaman,J., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

Submitted (23-NOV-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collymore,A., Cooke,A., Cooke,P., Corum,B., DeArellano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
 MacDonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
 Meldrim,J., Menueu,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
 Vassiliev,H., Venkataman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

TITLE

JOURNAL

Submitted (12-JUN-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Jun 12, 2003 this sequence version replaced gi:22381255.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L18049

Center clone name: 113_1_17

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 155596 bases at least Q40

Consensus quality: 156505 bases at least Q30

Consensus quality: 156946 bases at least Q20

Insert size: 182000; agarose-fp

Insert size: 156993; sum-of-contigs

Quality coverage: 10.7 in Q20 bases; agarose-fp

Quality coverage: 12.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

* 1 40918: contig of 40918 bp in length
 * 40919 41018: gap of 100 bp
 * 41019 41663: contig of 645 bp in length
 * 41664 41763: gap of 100 bp
 * 41764 43522: contig of 1759 bp in length
 * 43523 43622: gap of 100 bp
 * 43623 44624: contig of 1002 bp in length
 * 44625 44724: gap of 100 bp
 * 44725 47305: contig of 2581 bp in length
 * 47306 47405: gap of 100 bp
 * 47406 61096: contig of 13691 bp in length
 * 61097 61196: gap of 100 bp
 * 61197 81339: contig of 20143 bp in length
 * 81340 81439: gap of 100 bp
 * 81440 112412: contig of 30973 bp in length
 * 112413 112512: gap of 100 bp
 * 112513 157793: contig of 45241 bp in length.

FEATURES

source

1..157793
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP23-113117"
 /clone_lib="RPCI-23 Female Mouse BAC"

misc_feature

1..40918
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"

misc_feature

41019..41663

/note="assembly_fragment"

misc_feature

41764..43522

/note="assembly_fragment"

misc_feature

43623..44624

/note="assembly_fragment"

misc_feature

44725..47305

/note="assembly_fragment"

misc_feature

47406..61096

/note="assembly_fragment"

misc_feature

61197..81339

/note="assembly_fragment"

misc_feature

81440..112412

/note="assembly_fragment"

misc_feature

112513..157793

/note="assembly_fragment"

clone_end:T7
 vector_side:right"

BASE COUNT 42452 a 34156 c 35298 g 45105 t 812 others
 ORIGIN

```

/chrmosome="11"
/clone="RP11-183G22"
/clone_lib="RP11-11.1"
96048..96468
/note="Sequence from clone PCR only."
misc_feature
96070..96096
/note="Sequence from uni-directional dGTP big dye
terminator reads only."
BASE COUNT 41262 a 31518 c 29234 g 37259 t
ORIGIN
Query Match 0.8%; Score 22; DB 9; Length 139273;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1175 ATTCTGAAAGAAAGAACTGT 1196
|||||
DB 138156 ATTCTGAAAGAAAGAACTGT 138177

RESULT 35
AL928839 140119 bp DNA linear ROD 21-DEC-2002
LOCUS Mouse DNA sequence from clone RP23-298M2 on chromosome 2, complete
sequence.
ACCESSION AL928839
VERSION AL928839.10 GI:27368137
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 140119)
Griffiths, C.
Direct Submission
Submitted (20-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Dec 23, 2002 this sequence version replaced gi:27262731.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk

-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30). An attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Emi, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep RP23-298M2 is
from the RP11-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
Location/Qualifiers

```

```

source
1..140119
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-298M2"
/clone_lib="RP11-23"
BASE COUNT 42768 a 25589 c 25793 g 45969 t
ORIGIN
Query Match 0.8%; Score 22; DB 10; Length 140119;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 858 TCAAGAGATTAAAGTCAAGAAA 879
|||||
DB 114497 TCAAGAGATTAAAGTCAAGAAA 114518

RESULT 36
AC122459 156807 bp DNA linear ROD 19-OCT-2002
LOCUS Mus musculus chromosome 15 clone RP24-273F21, complete sequence.
DEFINITION AC122459
ACCESSION AC122459
VERSION AC122459.2 GI:24137602
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 156807)
McPherson, J.D. and Waterston, R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 156807)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 156807)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (19-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Oct 19, 2002 this sequence version replaced gi:21155920.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information
Center project name: M_B90273F21
-----
FEATURES
source
1..156807
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/clone="RP24-273F21"
BASE COUNT 42969 a 35824 c 36400 g 41614 t
ORIGIN
Query Match 0.8%; Score 22; DB 10; Length 156807;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1507 GGAGAACCGCGGAGATGGTG 1528
|||||
DB 153406 GGAGAACCGCGGAGATGGTG 153429

```



```

repeat_region complement(3519..3717)
repeat_region /rpt_family="AluSg/x"
repeat_region 5701..5804
repeat_region /rpt_family="MER5A"
repeat_region 5873..5971
repeat_region /rpt_family="L1M4"
repeat_region 7197..7559
repeat_region /rpt_family="MER20"
repeat_region 8700..8736
repeat_region /rpt_family="(TTG)n"
repeat_region 9223..9258
repeat_region /rpt_family="U2"
repeat_region 10121..10146
repeat_region /rpt_family="(TTTA)n"
repeat_region complement(10158..10386)
repeat_region /rpt_family="L1PAL0"
repeat_region complement(11193..11473)
repeat_region /rpt_family="AluJo"
repeat_region 11954..11993
repeat_region /rpt_family="(T)n"
repeat_region 12497..12631
repeat_region /rpt_family="MIR"
repeat_region 12633..12815
repeat_region /rpt_family="MSTC"
repeat_region 12954..13095
repeat_region /rpt_family="MSTC"
repeat_region 13871..14150
repeat_region /rpt_family="AluY"
repeat_region 14380..14563
repeat_region /rpt_family="MIR"
repeat_region complement(15192..15480)
repeat_region /rpt_family="AluSx"
repeat_region 15613..15671
repeat_region /rpt_family="MIR"
repeat_region complement(15994..16243)
repeat_region /rpt_family="L1PAL1"
repeat_region 16413..16688
repeat_region /rpt_family="AluY"
repeat_region 17979..18000
repeat_region /rpt_family="AT rich"
repeat_region complement(20594..20764)
repeat_region /rpt_family="L1PAL2"
STS 21925..22322
/standard_name="PZF_F/PZF_R"
/db_xref="GBD:214828"
22447..22548
/rpt_family="MIR"
23228..29686
/rpt_family="L1PAL0"
30266..30792
/rpt_family="MLTIF"
complement(31334..32052)
/rpt_family="MER44C"
32404..32529
/standard_name="DL251257"
/db_xref="DBSTS:8833"
32536..32640
/rpt_family="L2"
33140..33305
/rpt_family="MIR"
33343..33364
/rpt_family="(TTTA)n"
complement(33365..33444)
/rpt_family="L1PAL2"
33445..34137
/rpt_family="L1p"
35050..35147
/rpt_family="MIR"
36198..36627
/rpt_family="L1MA9"
complement(36628..36907)
/rpt_family="AluJo"
37345..39018

```

```

/rpt_family="L1PAL1"
complement(39019..39324)
/rpt_family="AluSc"
39325..39818
/rpt_family="L1PAL1"
39819..40218
Query Match C.8%; Score 22; DB 9; Length 127277;
Best Local Similarity 100.0%; Pred.No.7.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1178 CTGAAAGAAAGAACTGTTC A 1199
Db 9433 CTGAAAGAAAGAACTGTTC A 94315

RESULT 34
AL596275
LOCUS
DEFINITION Humar DNA sequence from clone Rp11-183G22 on chromosome 1, complete
sequence.
ACCESSION AL596275
VERSION AL596275.10 GI:17384503
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 139273)
Lawlor,S.
Direct Submission
Submitted (04-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Dec 5, 2001 this sequence version replaced gi:16973968.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: EM:, EMBL; SW:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
Rp11-183G22 is from the library RFL-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
Rp11-183G22. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone Rp11-183G22 is at 1 in this sequence.
The true left end of clone RP5-850015 is at 137274 in this
sequence. The true right end of clone Rp11-567C20 is at 88255 in
this sequence.

FEATURES
Location/Qualifiers
1..139273
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

```

repeat_region 31710..31761
/notes="2 copies 26 mer 96% conserved"
repeat_region 31736..31787
/notes="13 copies 4 mer tacc 100% conserved"
repeat_region 31788..32120
/notes="ALUSX repeat: matches 1..311 of consensus"
repeat_region 33330..33465
/notes="MIR repeat: matches 105..248 of consensus"
misc_feature complement(33467..33811)
/notes="match: GSS: Em:AQ114613"
repeat_region 34262..34356
/notes="MIR repeat: matches 37..131 of consensus"
repeat_region 35386..35589
Query Match 0.8%; Score 22; DB 9; Length 111799;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1175 ATTCTGGAAGAAGAACTGT 1196
|||||
Db 110919 ATTCTGGAAGAAGAACTGT 110898

RESULT 33
AC010175/c
LOCUS AC010175 127277 bp DNA linear PRI 31-DEC-2002
DEFINITION Homo sapiens 12 BAC RP11-118822 (Roswell Park Cancer Institute
Human BAC Library) complete sequence.
ACCESSION AC010175
VERSION AC010175.4 GI:6139075
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 127277)
AUTHORS Muzny,D.M., Adams,C., Bowtie,S., Brooks,A., Buhay,C., Bunac,C.,
Bodota,B., Bock,J., Bowler,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Dehazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,J.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
Holloway,C., Hosak,H., Issar,A., Jackson,L.E., Jackson,L., Jia,Y.,
Jones,M., Kelly,S., Kneitz,S., Kondejewski,N., Kong,Y., Kovar,C.,
Lau,S., Leal,B., Lee,E., Li,Z., Lichtarge,O., Liu,J., Liu,W.,
Logan,O., Lu,J., Lucier,R., Marondel,I., Martin,R., Martinez,C.,
McLeod,M.P., Mei,G., Merscher,S., Miller,A., Montgomery,K.T.,
Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N.,
Nguyen,S., Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L.,
Pu,L.L., Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J.,
Scherer,S., Shah,E., Shen,H., Shim,C., Simon,M., Sparks,A.,
Stamps,A., Sucgang,R., Tabot,P., Taylor,T., Vasquez,L., Vinson,R.,
Vo,O., Wahbah,M., Watlington,S., Weinstein,G., Weinstein,I.R.,
Williamson,A., Worley,K., Wren,J., Wrensford,G., Xiang,A.M.,
Yang,R., Yu,W., Zhou,X., Kucheriapati,R., Nelson,D. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 127277)
Worley,K.C.
Direct Submission
Submitted (15-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 127277)
Worley,K.C.
Direct Submission
Submitted (29-OCT-1999) Human Genome Sequencing Center, Department:
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 127277)
Worley,K.C.
Direct Submission

```

```

JOURNAL Submitted (16-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 127277)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (31-DEC-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Oct 29, 1999 this sequence version replaced gi:609161.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email:
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least 2 exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
c clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.
Location/Qualifiers
1..127277
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12p12-21.3-21.8"
/clone="RP11-118822"
1..132
/notes="overlaps bases 132 to 1 in AC007436."
/function="clone overlap"
complement(22..167)
/rpt_family="LIM3d"
complement(133..896)
/rpt_family="LIM3d"
960..1292
/rpt_family="L2"
complement(1342..1446)
/rpt_family="L1MA8"
1511..1942
/rpt_family="L2"
complement(2300..2470)
/rpt_family="MER58A"
3485..3518
/rpt_family="(TTTTC)n"

```



```

* 41340 41439: gap of 100 bp
* 41440 42110: contig of 671 bp in length
* 42111 42120: gap of 100 bp
* 42121 42863: contig of 653 bp in length
* 42864 42964: contig of 100 bp
* 42965 43614: contig of 651 bp in length
* 43615 43714: gap of 100 bp
* 43715 44362: contig of 648 bp in length
* 44363 44462: gap of 100 bp
* 44463 45008: contig of 546 bp in length
* 45009 45108: gap of 100 bp
* 45109 45756: contig of 648 bp in length
* 45757 45856: gap of 100 bp
* 45857 46512: contig of 656 bp in length
* 46513 46612: gap of 100 bp
* 46613 47287: contig of 675 bp in length
* 47288 47387: gap of 100 bp
* 47388 48042: contig of 655 bp in length
* 48043 48142: gap of 100 bp
* 48143 48808: contig of 666 bp in length
* 48809 48908: gap of 100 bp
* 48909 49576: contig of 668 bp in length
* 49577 49676: gap of 100 bp
* 49677 50333: contig of 657 bp in length
* 50334 50433: gap of 100 bp
* 50434 51111: contig of 678 bp in length
* 51112 51211: gap of 100 bp
* 51212 51870: contig of 659 bp in length.
Location/Qualifiers

```

Query Match 0.8%; Score 22; DB 2; Length 51870;

Best Local Similarity 100.0%; Pred. No. 8.4;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2164 GGAATTCAGGCCATCATGGTG 2185

Db 4897 GGAATTCAGGCCATCATGGTG 4876

RESULT 31

AC129491

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission
Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/SN/RepeatMasker.htm>.
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4475
Center clone name: 66_G14

* NOTE: This record contains 66 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

* 1 709: contig of 709 bp in length
* 710 809: gap of 100 bp
* 810 1530: contig of 721 bp in length
* 1531 1630: gap of 100 bp
* 1631 2337: contig of 707 bp in length
* 2338 2438: gap of 100 bp
* 2439 3145: contig of 707 bp in length
* 3146 3244: gap of 100 bp
* 3245 3946: contig of 702 bp in length
* 3947 4046: gap of 100 bp
* 4047 4755: contig of 709 bp in length
* 4756 4855: gap of 100 bp
* 4856 5543: contig of 688 bp in length
* 5544 6325: contig of 682 bp in length
* 6326 6425: gap of 100 bp
* 6426 7146: contig of 721 bp in length
* 7147 7246: gap of 100 bp
* 7247 7945: contig of 699 bp in length
* 7946 8045: gap of 100 bp
* 8046 8738: contig of 693 bp in length
* 8739 8838: gap of 100 bp
* 8839 9549: contig of 711 bp in length
* 9550 9649: gap of 100 bp
* 9650 10334: contig of 685 bp in length
* 10335 10434: gap of 100 bp
* 10435 11153: contig of 719 bp in length
* 11154 11253: gap of 100 bp
* 11254 11969: contig of 716 bp in length
* 11970 12069: gap of 100 bp
* 12070 12778: contig of 709 bp in length
* 12779 12878: gap of 100 bp
* 12879 13593: contig of 715 bp in length
* 13594 13693: gap of 100 bp
* 13694 14402: contig of 709 bp in length
* 14403 14502: gap of 100 bp
* 14503 15172: contig of 670 bp in length
* 15173 15272: gap of 100 bp
* 15273 15957: contig of 685 bp in length
* 15958 16057: gap of 100 bp
* 16058 16763: contig of 706 bp in length
* 16764 16863: gap of 100 bp
* 16864 17561: contig of 698 bp in length
* 17562 17661: gap of 100 bp
* 17662 18347: contig of 686 bp in length
* 18348 18447: gap of 100 bp
* 18448 19141: contig of 694 bp in length

```

McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
 Meneus, L., Mihova, T., Menges, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

Submitted (06-DEC-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information

Center project name: L19750
 Center clone name: 131_I_23

- * NOTE: This record contains 69 individual
- * sequencing reads that have not been assembled into
- * contigs. Runs of N are used to separate the reads
- * and the order in which they appear is completely
- * arbitrary. Low-pass sequence sampling is useful for
- * identifying clones that may be gene-rich and allows
- * overlap relationships among clones to be deduced.
- * However, it should not be assumed that this clone
- * will be sequenced to completion. In the event that
- * the record is updated, the accession number will
- * be preserved.

```

1
660: contig of 660 bp in length
661: gap of 100 bp
761: 1418: contig of 658 bp in length
1418: gap of 100 bp
1518: 2177: contig of 659 bp in length
2177: gap of 100 bp
2278: 2936: contig of 659 bp in length
2936: gap of 100 bp
3037: 3695: contig of 659 bp in length
3695: gap of 100 bp
3796: 4445: contig of 650 bp in length
4445: gap of 100 bp
4446: 4545: gap of 100 bp
4545: gap of 100 bp
4546: 5205: contig of 660 bp in length
5205: gap of 100 bp
5306: 5305: gap of 100 bp
5305: gap of 100 bp
5306: 6054: gap of 100 bp
6054: gap of 100 bp
6055: 6694: contig of 640 bp in length
6694: gap of 100 bp
6795: 6794: gap of 100 bp
6795: 7475: contig of 681 bp in length
7475: gap of 100 bp
7575: 8232: contig of 657 bp in length
8232: gap of 100 bp
8333: 8963: contig of 631 bp in length
8963: gap of 100 bp
9063: 9730: contig of 667 bp in length
9730: gap of 100 bp
9830: 10482: contig of 652 bp in length
10482: gap of 100 bp
10582: 11225: contig of 643 bp in length
11225: gap of 100 bp
11326: 11976: contig of 651 bp in length
11976: gap of 100 bp
12076: 12732: contig of 656 bp in length
12732: gap of 100 bp
12832: 13497: contig of 665 bp in length
13497: gap of 100 bp
13498: 13597: gap of 100 bp
13597: contig of 660 bp in length
14264: contig of 667 bp in length
14364: gap of 100 bp
15012: contig of 648 bp in length
15112: gap of 100 bp
15763: contig of 651 bp in length
15863: gap of 100 bp
16514: contig of 651 bp in length
16614: gap of 100 bp
17236: contig of 662 bp in length
17376: gap of 100 bp
18045: contig of 669 bp in length
18145: gap of 100 bp
18816: contig of 671 bp in length
18916: gap of 100 bp
19366: contig of 650 bp in length
19666: gap of 100 bp
20334: contig of 668 bp in length
20434: gap of 100 bp
21082: contig of 648 bp in length
21182: gap of 100 bp
21734: contig of 552 bp in length
21834: gap of 100 bp
22486: contig of 652 bp in length
22586: gap of 100 bp
23241: contig of 655 bp in length
23341: gap of 100 bp
23991: contig of 650 bp in length
24091: gap of 100 bp
24742: contig of 651 bp in length
24842: gap of 100 bp
25507: contig of 665 bp in length
25607: gap of 100 bp
26261: contig of 654 bp in length
26361: gap of 100 bp
27033: contig of 652 bp in length
27133: gap of 100 bp
27772: contig of 659 bp in length
27872: gap of 100 bp
28540: contig of 668 bp in length
28640: gap of 100 bp
29290: contig of 650 bp in length
29390: gap of 100 bp
30037: contig of 647 bp in length
30137: gap of 100 bp
30790: contig of 653 bp in length
30890: gap of 100 bp
31531: contig of 641 bp in length
31631: gap of 100 bp
32282: contig of 651 bp in length
32382: gap of 100 bp
33036: contig of 654 bp in length
33136: gap of 100 bp
33799: contig of 663 bp in length
33899: gap of 100 bp
34556: contig of 657 bp in length
34656: gap of 100 bp
35301: contig of 645 bp in length
35401: gap of 100 bp
36054: contig of 653 bp in length
36154: gap of 100 bp
36818: contig of 664 bp in length
36918: gap of 100 bp
37572: contig of 654 bp in length
37672: gap of 100 bp
38318: contig of 646 bp in length
38418: gap of 100 bp
39072: contig of 654 bp in length
39172: gap of 100 bp
39835: contig of 663 bp in length
39935: gap of 100 bp
40579: contig of 644 bp in length
40679: gap of 100 bp
41339: contig of 660 bp in length

```



```

RESULT 27
AF272662      5551 bp      mRNA      linear      ROD 11-SEP-2000
LOCUS         Rattus norvegicus alpha 1 type V collagen mRNA, complete cds.
DEFINITION    AF272662
ACCESSION     AF272662
VERSION       AF272662.1 GI:8568095
KEYWORDS      Rattus norvegicus (Norway rat)
SOURCE        Rattus norvegicus
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
REFERENCE     1 (bases 1 to 5551)
Chernousov,M.A., Rothblum,K., Tyler,W.A., Stahl,R.C. and Carey,D.J.
Schwann cells synthesize type V collagen that contains a novel
alpha 4 chain. Molecular cloning, biochemical characterization, and
high affinity heparin binding of alpha 4(V) collagen
J. Biol. Chem. 275 (36), 28208-28215 (2000)
20428740
MEDLINE       10852920
PUBMED        10852920
REFERENCE     2 (bases 1 to 5551)
Chernousov,M.A., Rothblum,K.N., Tyler,W.A., Stahl,R.C. and
Carey,D.J.
Direct Submission
Submitted (26-MAY-2000) Weis Center for Research, Penn State
College of Medicine, 100 N. Academy Ave., Danville, PA 17822, USA
Location/Qualifiers
1. .5551
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague Dawley"
/db_xref="taxon:10116"
/cell_type="Schwann"
6. .5528
/notes="collagen alpha chain"
/codon_start=1
/product="alpha 1 type V collagen"
/protein_id="AAF76433.1"
/db_xref="GI:8568096"
/translation="MDVHTRWKAPRPGAPLLSPILLLLLLLLWAPPRAAQTDLLE
MLDFHNLPSGVTKTTCFATRRSRGPDVAYRVSKDQLSMPTKOLYPSDFEDFSI
LTTVAKKGSQAFVSVYNEQGIQLQLELGRSPVLYEDHTGKPGSEYPLFPGINL
SDGKWHRIASVYKXNVTLIDCKKIKFLNRGHPHIDVNGIMFGSRILDEIFE
GDIQQLFVSDHRAAYDYCEHVSPTDTPVPTQSDPNPDEYVEGEETVYVEYP
YVEDPDGKEAPAPQKVEAEARETTEVEEQTPQSEAPTVPETSDTAGKEDNPIG
DIDYVTDYTTSPYEDGYGEGVENPQPNPDSGABIPSTSVTSNSNPASPE
EGKDDLGGEFTETIKNLENYDYPIFDPSDSNVSPSEIGFPMQANQDTIYEGIGP
RGEKQKGEPAIIEPQMLIEGPPGEPAGLPFGPPTGTQMGDPGERGEPGRPL
PGADGLPGPPTMLMPLFRFGGSDAGSKGPMVSAESAQAIILOQALRALRGPAGM
GLTRGPGMPGPGSGGLGKEBDMGPGQGRVQVGPQTKPGRGRAGSJCARGMPG
OTGPKGDRGDFDGLAGLPGKGRHGDPSGPGPLGDDGERGDEGVGRGLPGPBP
RGLLGPKPGPGPGVGTGMDQPGQKNGVQCEGPPGCGQPGPGLPGPOGA1
GPPGKEGPKLQGLPMQADGPPHGPKEGPPGEGKGGGPPGPGPTGYGPRGVK
ADGIRGLTKGKBGSDGFPFGKDMGIKDRGEIGPPGPRGEDGPEGKPGKGNPD
PGPLTGTGKGLGVPLGPGYVQGRQPKGSIGFPGFGANGKGRGTGPKGPRGR
GPTGPRGERPRGITGKPGPKGNSGDDGAGPPGPRGPNQGTGTPGKPGPPPG
PGAGKEGPKDGPAGLQKQDGPGLRPFGRDGLPGVAGLGLKSGEPGPG2PAPA
GSGFERGPAAGAPIGIPGRPGQPGPPGAGKEGVPEEGPQGPAGRDGLQGVGLPG
PAGVPVPGEDGKBGSDGIFGPGKSGDKDGEQFPPTGPTQPTGQPGSGADGFP
RQGGAPQPGKSGSGFPGVGLQGLPGPPGKEGTGVDGQMGPPGPPGPRGPG
GARGADGPPGPGIGNPAGVGEKGEAGEGPGLPGEGLPGKGERGKEAGEGFS
AAGPGKPGPDGPGKSGPVGFPDGPPEPGACQDGPBGDKDGDGPGPGTGS
PGTGPBPGSGPGKPGPQAPGEPGQKGNKAGLGPCKGTGLPGQAGKPG
GPGLKGIQPVGVQSLGSPGPDGPPGFWMPGPPGLGLKSDSGPKGKHPLGLIG
PPGEQKEDRGLPFGSGSGVEQGITGSPGLPGPPGLPGPPGPKAGSGSGF
TKPGKAGHGLPFGPPGSEVITQPLIQASRTNRINDASQLDDAGSEYVDYADG
EETFGSINLSKLEIEQMKRPLGTQNPARTCKQLQCHPDPDGEVYVDPNQCSDS
KVVYCNATAGSTVCPDKSEKARTISWPNKSGMSFSEFKGLLSVDAEENPVG
VWOMTELRLLSAAQONITVYQSVAMODATGSDKAIREFLGSNDSEMSYDNPYI
RALVDGCAKTKGQKTVLEIDTPKVEQVFIVDIMFTDGEASQKGFVEVGPACFLQ"

```

```

BASE COUNT    1214 a    1645 c    1773 g    919 t
ORIGIN
Query Matche      0.8%; Score 22; DB 10; Length 5551;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2702 GCACAGAGAGGCTCTCCAGGAGA 2723
Dc 2308 GCAAAGAAGGCTCTCCAGGAGA 2329
RESULT 28
CRUPA1V
LOCUS         CRUPA1V      6114 bp      mRNA      linear      ROD 27-APR-1993
DEFINITION    Chinese hamster pro-alpha-1 (V) collagen mRNA, complete cds.
ACCESSION     M76730
VERSION       M76730.1 GI:191150
KEYWORDS      alpha-1 type V collagen.
SOURCE        Cricetus longicaudatus (Long-tailed hamster)
ORGANISM      Cricetus longicaudatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetus
1 (bases 1 to 6114)
Greenspan,D.S., Cheng,W. and Hoffman,G.G.
The pro-alpha 1(V) collagen chain: Complete primary structure,
distribution of expression, and comparison with the pro-alpha 1(XI)
collagen chain
J. Biol. Chem. 266 (36), 24727-24733 (1991)
92105142
PUBMED        172213
COMMENT       Original source text: Cricetus longicaudatus male adult cDNA to
mRNA.
FEATURES      Location/Qualifiers
1. .6114
/organism="Cricetus longicaudatus"
/mol_type="mRNA"
/db_xref="taxon:10030"
/sex="male"
/dev_stage="adult"
1. .6114
/genes="pro-alpha-1 (V) collagen"
33. .5555
/genes="pro-alpha-1 (V) collagen"
/codon_start=1
/product="pro-alpha-1 type V collagen"
/protein_id="AA337002.1"
/db_xref="GI:191151"
/translation="MDVHTRWKDRLPVGAAPVPLLLLLLLWAPPQRAAQTDLLE
MLDFHNLPSGVTKTTCFATRRSRGPDVAYRVSKDQLSMPTKOLYPSDFEDFSI
LTTVAKKGSQAFVSVYNEQGIQQLQLELGRSPVLYEDHTGKPGSEYPLFPGINL
SDGKWHRIASVYKXNVTLIDCKKIKFLNRSDHPHIDVNGIMFGSRILDEIFE
GDIQQLFVSDHRAAYDYCEHVSPTDTPVPTQSDPNPDEYVEGEETVYVEYP
YVEDPDGKEAPATQKVEAEARETTEVEEQTPQSEAPTVPETSDTAGKEDDPFG
DIDYVTDYTTTPYEDGYGEGVENPQPNPDSGAEVPTSTITSNTSNFAPSPG
EDKDDLGGEFTETIKNLENYDYPIFDPSDSNVSPSEIGFPMQANQDTIYEGIGP
RGEKQKGEPAIIEPQMLIEGPPGEPAGLPFGPPTGTQMGDPGERGEPGRPL
PGADGLPGPPTMLMPLFRFGGSDAGSKGPMVSAESAQAIILOQALRALRGPAGM
GLTRGPGMPGPGSGGLGKEBDMGPGQGRVQVGPQTKPGRGRAGSJCARGMPG
OTGPKGDRGDFDGLAGLPGKGRHGDPSGPGPLGDDGERGDEGVGRGLPGPBP
RGLLGPKPGPGPGVGTGMDQPGQKNGVQCEGPPGCGQPGPGLPGPOGA1
GPPGKEGPKLQGLPMQADGPPHGPKEGPPGEGKGGGPPGPGPTGYGPRGVK
ADGIRGLTKGKBGSDGFPFGKDMGIKDRGEIGPPGPRGEDGPEGKPGKGNPD
PGPLTGTGKGLGVPLGPGYVQGRQPKGSIGFPGFGANGKGRGTGPKGPRGR
GPTGPRGERPRGITGKPGPKGNSGDDGAGPPGPRGPNQGTGTPGKPGPPPG
PGAGKEGPKDGPAGLQKQDGPGLRPFGRDGLPGVAGLGLKSGEPGPG2PAPA
GSGFERGPAAGAPIGIPGRPGQPGPPGAGKEGVPEEGPQGPAGRDGLQGVGLPG
PAGVPVPGEDGKBGSDGIFGPGKSGDKDGEQFPPTGPTQPTGQPGSGADGFP
RQGGAPQPGKSGSGFPGVGLQGLPGPPGKEGTGVDGQMGPPGPPGPRGPG
GARGADGPPGPGIGNPAGVGEKGEAGEGPGLPGEGLPGKGERGKEAGEGFS
AAGPGKPGPDGPGKSGPVGFPDGPPEPGACQDGPBGDKDGDGPGPGTGS
PGTGPBPGSGPGKPGPQAPGEPGQKGNKAGLGPCKGTGLPGQAGKPG
GPGLKGIQPVGVQSLGSPGPDGPPGFWMPGPPGLGLKSDSGPKGKHPLGLIG
PPGEQKEDRGLPFGSGSGVEQGITGSPGLPGPPGLPGPPGPKAGSGSGF
TKPGKAGHGLPFGPPGSEVITQPLIQASRTNRINDASQLDDAGSEYVDYADG
EETFGSINLSKLEIEQMKRPLGTQNPARTCKQLQCHPDPDGEVYVDPNQCSDS
KVVYCNATAGSTVCPDKSEKARTISWPNKSGMSFSEFKGLLSVDAEENPVG
VWOMTELRLLSAAQONITVYQSVAMODATGSDKAIREFLGSNDSEMSYDNPYI
RALVDGCAKTKGQKTVLEIDTPKVEQVFIVDIMFTDGEASQKGFVEVGPACFLQ"

```

CDS


```

/locus_tag="OSJNBb0079E01.10"
/codon_start=1
/product="putative cytochrome 450"
/protein_id="AAP52886.1"
/db_xref="GI:31431055"
/translation="MEVAAMEISTSLTTLTVALSVIVCYALVFRAGKARAPLEPLPFG
PRGVPVLGNLPOLGKTKQTLEHMTKQYGLPIRLRFSGSDVVVAGSAPVAAQFLRTHD
ANFSRPNKSGGHEMAYNGRDVVPFGYGRWRMRKICAVNLFSARALDDLRAPERE
AVLMVRSIAESAAPGSSSSPAAVLVKGVNCTTNALSRAAVGRVFAAGAGEGARF
KEILVEVGVGVNAGFVPALRWLDQGVVARMKKLRRFPDDMMALIAERRAGSL
LKPTDSEEGKDLGULLAMVOEQELWAAGEDDRIITDEIKALINLUFVAGTUTIST-
VENTMAELIRHPDILHQAEEEDVVGDRILLSEDSLSTFFHAIKETFLRHPSTP
LSLPRMASEECBIAGYRIPKGAELLVNMGARDAIPWDELYKPSRFLPFGGTHDV
DVKGNDFGLIPFAGGRICPKAGSLMGLRMVTMTAATLVHAFDWQLPADQTPDKLANDEA
FTLLQRAEPLVHPVPLLPSPYNIA"
/complement(<35810..>36529)
/locus_tag="OSJNBb0079E01.9"
/note="hypothetical protein"
/complement(join(<35810..36283,36485..>36529))
/locus_tag="OSJNBb0079E01.9"
/complement(join(35810..36283,36485..36529))
/locus_tag="OSJNBb0079E01.9"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAP52887.1"
/db_xref="GI:31431056"
/translation="MEISGALVAOMRWLGGVAVSGKEVPIAVTRDLRQGGGRNSGVGR
IVPAVPAPESIVEGRRSRSGAGPLGGVWLFATWDGAGPAVRMRVPVASKEDACG
RSEEAAGNRGESARRGFARVIAVNRNCNRGGQMRDKIRVDNRKTNENRGCDFT
RIARHNSFLY"
<43732..>44121
/locus_tag="OSJNBb0079E01.8"
/note="hypothetical protein"
<43732..>44121
/locus_tag="OSJNBb0079E01.8"
43732..44121
/locus_tag="OSJNBb0079E01.8"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAP52888.1"
/db_xref="GI:31431057"
/translation="WATRDSGGDAQLRHPAPAAAARLQOQRRRTNAVAQQGSAAVAV
AAARQRQAAVHGTTHSSGALGAGARVYRGDTTITRQLGHAVRVNAKRSHLHKE
GLEKISHRKQRQCGFTAVTVLRG"
47723..>48577
/locus_tag="OSJNBb0079E01.7"
/note="hypothetical protein"
join(<47723..47818,48329..>48577)
/locus_tag="OSJNBb0079E01.7"
join(47723..47818,48329..48577)
/locus_tag="OSJNBb0079E01.7"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAP52889.1"
/db_xref="GI:31431058"
/translation="WALTGARSKENMKNSLRFHSYTLPSVPVAAQQHILAFSRFGX
EKIRIELSWENRQQRKKGEEDERSVCLPYEPGLGARPGGGMGLGEBRSYVTGGKSS
WVHGTDKRLN"
<59545..>63413
/locus_tag="OSJNBb0079E01.6"
/note="Similar to sp Q05654 RDPO_SCHPO Retrotransposable
element Tf2 155 kba protein E value 2e 43; contains pfam
domain: rve Integrase core domain e value 1 2e 36;
contains pfam domain: rvt Reverse transcriptase (RNA
dependent DNA e value 5 8e 20; contains pf"
join(<59545..60258,60292..50786,60883..61408,61536..61676,
61918..>63413)
/locus_tag="OSJNBb0079E01.6"
join(59545..60258,60292..60786,60883..61408,61536..61676,
61918..63413)
/locus_tag="OSJNBb0079E01.6"
/codon_start=1
/product="putative retroelement"

```

```

/protein_id="AAP52890.1"
/db_xref="GI:31431059"

Query Match      0.8%; Score 23; DB 8; Length 304;10;
Best Local Similarity 100.0%; Pred. No. 2;1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1647 AAAAGGTGAAAGGGGAATGCTG 1669
|||||.....|
DB 48755 AAAAGGTGAAAGGGGAATGCTG 48777

RESULT 26
AY158896      2733 bp      mRNA      linear      ROD 03-JUN-2003
LOCUS      AY158896
DEFINITION      Rattus norvegicus alpha 1 type XXIII collagen (Col23a1) mRNA,
complete cds.
ACCESSION      AY158896
VERSION      AY158896.1 GI:29373915
KEYWORDS
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Sutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE      1 (bases 1 to 2733)
            Banyard,J., Bao,L. and Zetter,B.R.
            Type XXIII Collagen, a New Transmembrane Collagen Identified in
            Metastatic Tumor Cells
            C. Biol. Chem. 278 (23), 20989-20994 (2003)
JOURNAL      22662249
MEDLINE      12644459
REFERENCE      2 (bases 1 to 2733)
            Banyard,J., Bao,L. and Zetter,B.R.
            Direct Submission
TITLE      Submitted (04-OCT-2002) Department of Surgery, Children's Hospital,
            300 Longwood Ave., Boston, MA 02115, USA
JOURNAL      Location/Qualifiers
FEATURES
            1..2733
                /organism="Rattus norvegicus"
                /mol_type="mRNA"
                /strain="Copenhagen"
                /db_xref="taxon:10116"
                /coll_line="A16.1"
                /cell_type="Dunning R-3327 prostate adenocarcinoma"
            6
                /gene="Col23a1"
                /gene="Col23a1"
                /gene="Col23a1"
                /codon_start=1
                /product="alpha 1 type XXIII collagen"
                /protein_id="AA018362.1"
                /db_xref="GI:29373916"
                /translation="MGERRAAGGGTQDPGAGCCARALGALCLLSVGSATACLLLG
                AQAALHGVAALEQRELLRAGSGALAAWETHLERLLREKLDGVAKRTVREAP
                SECICPPGRRGKRRGDPGPPQSGRDGYPGLGLDGKPLGPKGKGAQGF
                GPCACQCGAAGPPPGAPGAPPGTGDGPRGAQCPGPGESQDGMGPMG
                PPGKPGCTPGKKGGDGI PSQPLPGPPKPGCGVPGQGTGVDGAPGLKGEFGH
                PGTDGALGRGPPGLKGEQGDVWIDYDGRLLDALKGPFGGAPGPGPGI PGAKGELG
                LQAPGIDSEKPGKPGKPGPPGPPGAGPKGTGEMGLSGPADQPKGKESASDHL
                QESLAQIIVPPGPPGPPGPMGLQIGPKGLDGAKEGKASGERGPHGLPVPVP
                PGLIGLPGTKGKGRGPEGLDGFPGPRGKGRGDRSERGKGVFGRKGVQKGE
                GPFGLDQPCVPDPGLVPFCWHK"
BASE COUNT      598 a      808 c      904 g      423 t
ORIGIN
Query Match      0.8%; Score 22; DB 10; Length 2733;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2499 AGGCCAGAGGGTCCAGAGGA 2520
|||||.....|
DB 725 AGGCCAGAGGGTCCAGAGGA 746

```

[illegible]

Paro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Katat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehocsky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,I., Mihova,T., Mlenga,V., Murphy,I., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (06-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 188326)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barns,N., Bastien,V., Bloom,I., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Katat,A., Karatas,A., Kells,C., Lakocque,K., Lamazares,R., Landers,T., Lehocsky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,I., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (31-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 31, 2002 this sequence version replaced gi:20455631.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RX/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22138
Center clone name: 713_M15

FEATURES

source
1. .188326
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-713M15"
/clone_lib="RP11-11 Human Male BAC"
complement(138. .370)
/rpt_family="MIR"
repeat_region
complement(540. .843)
repeat_region
/rpt_family="AluJb"
repeat_region
complement(1342. .1729)

/rpt_family="L1MC4"
1784. .2026
/rpt_family="L1MC4"
2092. .2458
/rpt_family="L1MC4"
2480. .2532
/rpt_family="MERSA"
complement(3120. .3522)
/rpt_family="MLT1A1"
3599. .3683
/rpt_family="L2"
3684. .3923
/rpt_family="MIR"
4408. .4557
/rpt_family="MUTLK"
complement(5098. .5184)
/rpt_family="MIR3"
complement(5394. .5647)
/rpt_family="L2"
7198. .7700
/rpt_family="L2"
7974. .8128
/rpt_family="MIR3"
8287. .8466
/rpt_family="L1PA15"
complement(8467. .8774)
/rpt_family="AluSx"
8775. .8797
/rpt_family="L1PA15"
9069. .9274
/rpt_family="MIR"
11640. .11666
/rpt_family="CGG)n"
11845. .11908
/rpt_family="GC_rich"
12223. .12246
/rpt_family="GC_rich"
complement(14375. .14433)
/rpt_family="MERSA"
14788. .14879
/rpt_family="MIR"
complement(15115. .15269)
/rpt_family="MIR"
15270. .15571
/rpt_family="AluSg"
complement(15572. .15658)
/rpt_family="MIR"
complement(16044. .16174)
/rpt_family="MER91C"
16693. .16845
/rpt_family="MIR3"
complement(16948. .17146)
/rpt_family="MER20B"
complement(17166. .17328)
/rpt_family="MER20B"
complement(17422. .17618)
/rpt_family="MIR"
complement(17619. .17720)
/rpt_family="MIR"
complement(17749. .18274)
/rpt_family="MLT1D"
19131. .19174
/rpt_family="L2"
19218. .19322
/rpt_family="GA-rich"
20495. .20710
/rpt_family="MIR"
21664. .21765
/rpt_family="L1MED"
complement(21766. .21862)
/rpt_family="MER104"
22027. .22086
/rpt_family="AT_rich"

[illegible]

```

repeat_region 29327..29456
/evdence=not_experimental
/rpt_family="L16C"
repeat_region 29510..29865
/evdence=not_experimental
/rpt_family="HAL1"
repeat_region complement(29895..30257)
/evdence=not_experimental
/rpt_family="MER119"
repeat_region 30514..30739
/evdence=not_experimental
/rpt_family="HAL1"
repeat_region 30742..31096
/evdence=not_experimental
/rpt_family="HAL1"
repeat_region 31126..31174
/evdence=not_experimental
/rpt_family="AT_rich"
repeat_region 31572..31631
/evdence=not_experimental
/rpt_family="AT_rich"
repeat_region 31796..31892
/evdence=not_experimental

Query Match 0.84; Score 23; DB 9; Length 151880;
Best Local Similarity 100.0%; Pred. No. 2,2; Mismatches 0; Indels 0; Gaps 0;
Matches 23; Conservative 0;

OY 1450 CCAGTGTTCCAGGATCTCCAGG 1472
|||||
DB 36464 CCAGTGTTCCAGGATCTCCAGG 36442

RESULT 23
AC106013/C
LOCUS AC106013 166093 bp DNA linear PRI 02-MAR-2002
DEFINITION Homo sapiens chromosome 8, clone RP11-1109L18, complete sequence.
ACCESSION AC106013
VERSION AC106013.2 GI:18543063
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 166093)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Minova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Olivier,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Roman,J., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
2 (bases 1 to 166093)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Minova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Olivier,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Roman,J., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Submitted (11-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 166093)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Minova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivier,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (02-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 6, 2002 this sequence version replaced gi:18129519.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WISR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22151
Center clone name: 1109_L18
-----
Location/Qualifiers
1. 166093
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-1109L18"
/clone_lib="RP11-11 Human Male BAC"
repeat_region 1..103
/rpt_family="AluSg/x"
complement(550..747)
/rpt_family="MIR3"
repeat_region 1559..1836
/rpt_family="AluSx"
repeat_region 2812..2909
/rpt_family="L2"
repeat_region 2913..3032
/rpt_family="MIR"
repeat_region 3123..3423
/rpt_family="AluSx"
complement(3449..3646)
/rpt_family="MIR"
repeat_region 4330..4742
/rpt_family="MLT1B"
repeat_region complement(4899..5264)
/rpt_family="THE1C"
repeat_region 5877..5929
/rpt_family="CT-rich"
repeat_region 6015..6149
/rpt_family="MIR3"
complement(6127..6218)
/rpt_family="MIR"

```


* 41180 contig of 723 bp in length
 * 42002 gap of 100 bp
 * 42003 contig of 711 bp in length
 * 42714 gap of 100 bp
 * 42814 contig of 702 bp in length
 * 43515 gap of 100 bp
 * 43616 contig of 699 bp in length
 * 44314 gap of 100 bp
 * 44414 gap of 100 bp
 * 44415 contig of 708 bp in length
 * 45123 gap of 100 bp
 * 45223 contig of 721 bp in length
 * 45944 gap of 100 bp
 * 46044 contig of 681 bp in length
 * 46824 gap of 100 bp
 * 46825 contig of 728 bp in length
 * 47553 gap of 100 bp
 * 47653 contig of 728 bp in length
 * 48381 gap of 100 bp
 * 48481 contig of 718 bp in length
 * 49199 gap of 100 bp
 * 49299 contig of 722 bp in length
 * 50021 gap of 100 bp
 * 50131 contig of 716 bp in length
 * 50837 gap of 100 bp
 * 50937 contig of 719 bp in length
 * 51655 gap of 100 bp
 * 51755 gap of 100 bp
 * 52470 contig of 714 bp in length
 * 52569 gap of 100 bp
 * 53370 contig of 701 bp in length
 * 53371 gap of 100 bp
 * 54076 contig of 706 bp in length
 * 54176 gap of 100 bp
 * 54857 contig of 681 bp in length
 * 54858 gap of 100 bp
 * 54957 gap of 100 bp
 * 55643 contig of 686 bp in length
 * 55743 gap of 100 bp
 * 55744 contig of 717 bp in length

Query Match 0.8%; Score 23; DB 2; Length 67884;

Best Local Similarity 100.0%; Pred. No. 2,3;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1450 CCAGGTGTCAGGATCTCCAGG 1472

DB 25961 CCAGGTGTCAGGATCTCCAGG 25939

RESULT 21

AC119149/c

LOCUS

DEFINITION Oryza sativa (japonica cultivar-group) chromosome 10 clone

AC119149

ACCESSION AC119149.2 GI:21306633

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
 NY 11724, USA
 3 (bases 1 to 132870)
 McCombie, W.R.
 Direct Submission
 Submitted (01-JUN-2002) Lita Annenberg Hazen Genome Center, Cold
 Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
 NY 11724, USA
 4 (bases 1 to 132870)
 Palmer, L.E., de la Bastide, M., Spiegel, L., Preston, R., Ferraro, K.,
 Kuit, K., Nascimento, L., Zucavern, T., Ballia, V., Bell, M., Baker, C.,
 Miller, B., Katzenberger, F., Muller, S., King, L., Sullivan, P.,
 Yang, C., Dike, S., O'Shaughnessy, A., Dedhia, N., and McCombie, W.R.
 Direct Submission
 Submitted (02-JUL-2002) Lita Annenberg Hazen Genome Center, Cold
 Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY
 11724, USA
 Oryza sativa (japonica cultivar-group) chromosome 10 clone
 OSJNB0079E01, complete sequence.

5 (bases 1 to 132870)
 McCombie, W.R.
 Direct Submission

Submitted (14-JUL-2002) Genome Research Center, Cold Spring Harbor
 Laboratory, 500 Sunnyside Blvd, Woodbury, NY 11797, USA

On Jun 1, 2002 this sequence version replaced GI:20303540.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The nucleotide

sequence of this BAC clone was generated by combining Syngenta and
 Cold Spring Harbor Laboratory Genome Center sequencing data.

Clone OSJNB0079E01 overlaps clone OSJNB0053D03 (AC021892) from
 base 1 to base 163777. The overlap is from base 163777 to base

199113 on OSJNB0053D03. Clone OSJNB0079E01 also overlaps clone
 OSJNB0036B06 (AC051634) from base 121446 to base 132870. The

overlap is from base 1 to base 11425 on OSJNB0036B06.

Location/Qualifiers

1..132870

/organism="Oryza sativa (japonica cultivar-group);"

/mol_type="genomic DNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/chromosome="10"

/clone="OSJNB0079E01"

/clone_lib="EccRI"

/complement(2679..2700)

/note="Low complexity AT-rich"

2738..2914

/note="Simple repeat (CCGG)n"

/complement(6746..6946)

/note="Similar to unknown retrotransposon"

/complement(6942..7610)

/note="Similar to unknown retrotransposon"

/complement(9396..9535)

/note="Similar to putative transposon, CACTA, En/Spm-like"

/complement(10340..10384)

/note="Low complexity GC-rich"

12175..12204

/note="Simple repeat (GA)n"

12544..12867

/gene="OSJNB0079E01.2"

/note="Hypothetical protein"

12544..12867

/gene="OSJNB0079E01.2"

/codon_start=1

/product="Hypothetical protein"

/protein_id="RAM74386.1"

/db_xref="GI:21672024"

/translation="MAAKVTDQCKLKLKLVKEDSDRKESEKRESESEKRIEF

NALKVWESHFTVVEKRVDELQGSWELQKVEKLOIVVQASREPFAPTSPTGALS

Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,
 Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T.,
 Lehoczyk, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
 McPheeters, R., Melidrim, J., Meneus, L., Minova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
 Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
 Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Travers, B., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, M., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, N.

Direct Submission

Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L11533

Center clone name: 567_N_24

- * NOTE: This record contains 84 individual
- * sequencing reads that have not been assembled into
- * contigs. Runs of N are used to separate the reads
- * and the order in which they appear is completely
- * arbitrary. Low-pass sequence sampling is useful for
- * identifying clones that may be gene-rich and allows
- * overlap relationships among clones to be deduced.
- * However, it should not be assumed that this clone
- * will be sequenced to completion. In the event that
- * the record is updated, the accession number will
- * be preserved.

1 584: contig of 684 bp in length
 685 784: gap of 100 bp
 785 1507: contig of 723 bp in length
 1508 1607: gap of 100 bp
 1608 2320: contig of 713 bp in length
 2321 2420: gap of 100 bp
 2421 3131: contig of 711 bp in length
 3132 3231: gap of 100 bp
 3232 3907: contig of 676 bp in length
 3908 4007: gap of 100 bp
 4008 4711: contig of 704 bp in length
 4712 4811: gap of 100 bp
 4812 5512: contig of 701 bp in length
 5513 5613: contig of 100 bp
 5613 6316: contig of 704 bp in length
 6317 6416: gap of 100 bp
 6417 7135: contig of 719 bp in length
 7136 7235: gap of 100 bp
 7236 7934: contig of 719 bp in length
 7935 8054: gap of 100 bp
 8055 8776: contig of 722 bp in length
 8777 9593: contig of 717 bp in length
 9594 10411: gap of 100 bp
 10412 10511: contig of 718 bp in length
 10512 11244: contig of 713 bp in length
 11245 11324: gap of 100 bp
 11325 12028: contig of 704 bp in length

12029 12128: gap of 100 bp
 12129 12806: contig of 678 bp in length
 12807 12906: gap of 100 bp
 12907 13610: contig of 704 bp in length
 13611 13710: gap of 100 bp
 13711 14336: contig of 726 bp in length
 14337 14536: gap of 100 bp
 14537 15263: contig of 727 bp in length
 15264 15363: gap of 100 bp
 15364 16080: contig of 717 bp in length
 16081 16180: gap of 100 bp
 16181 16889: contig of 709 bp in length
 16890 16989: gap of 100 bp
 16990 17088: contig of 719 bp in length
 17089 17808: gap of 100 bp
 17809 18531: contig of 723 bp in length
 18532 18631: gap of 100 bp
 18632 19323: contig of 692 bp in length
 19324 19423: gap of 100 bp
 19424 20130: contig of 707 bp in length
 20131 20230: gap of 100 bp
 20231 20925: contig of 695 bp in length
 20926 21025: gap of 100 bp
 21027 21728: contig of 703 bp in length
 21729 21828: gap of 100 bp
 21829 22521: contig of 693 bp in length
 22522 22621: gap of 100 bp
 22622 23114: contig of 693 bp in length
 23115 23414: gap of 100 bp
 23415 24146: contig of 732 bp in length
 24147 24246: gap of 100 bp
 24247 24961: contig of 715 bp in length
 24962 25061: gap of 100 bp
 25062 25771: contig of 710 bp in length
 25772 25871: gap of 100 bp
 25872 26570: contig of 699 bp in length
 26571 26670: gap of 100 bp
 26671 27388: contig of 718 bp in length
 27389 27488: gap of 100 bp
 27489 28192: contig of 704 bp in length
 28193 28292: gap of 100 bp
 28293 28999: contig of 707 bp in length
 29000 29099: gap of 100 bp
 29100 29807: contig of 708 bp in length
 29808 29907: gap of 130 bp
 29908 30600: contig of 693 bp in length
 30601 30700: gap of 100 bp
 30701 31403: contig of 703 bp in length
 31404 31503: gap of 100 bp
 31504 32219: contig of 716 bp in length
 32220 32319: gap of 100 bp
 32320 33034: contig of 715 bp in length
 33035 33134: gap of 100 bp
 33135 33855: contig of 721 bp in length
 33856 33955: gap of 100 bp
 33956 34658: contig of 703 bp in length
 34659 34759: gap of 100 bp
 34759 35461: contig of 703 bp in length
 35462 35561: gap of 100 bp
 35562 36265: contig of 704 bp in length
 36266 36365: gap of 100 bp
 36366 37082: contig of 717 bp in length
 37083 37182: gap of 100 bp
 37183 37968: contig of 686 bp in length
 37969 38673: gap of 100 bp
 38674 38773: contig of 705 bp in length
 38774 39462: contig of 689 bp in length
 39463 39562: gap of 100 bp
 39563 40254: contig of 692 bp in length
 40255 40354: gap of 100 bp
 40355 41079: contig of 725 bp in length
 41080 41179: gap of 100 bp

TITLE JOURNAL

COMMENT

```

repeat_region 8435. .8737
/note="AluX repeat: matches 1. .298 of consensus"
repeat_region 8738. .9144
/note="L1M2 repeat: matches 4193. .4614 of consensus"
repeat_region 9339. .11100
/note="MER66-internal repeat: matches 4895. .6676 of consensus"
repeat_region 11101. .11496
/note="MER66B repeat: matches 1. .392 of consensus"
repeat_region 11510. .12032
/note="L1M2 repeat: matches 4604. .5114 of consensus"
repeat_region 12033. .12338
/note="AluX repeat: matches 1. .306 of consensus"
repeat_region 12339. .13410
/note="L1M2 repeat: matches 5114. .6307 of consensus"
repeat_region 13633. .13852
/note="AluY repeat: matches 87. .307 of consensus"
repeat_region 13896. .15155
/note="TIGER2 repeat: matches 1266. .2718 of consensus"
repeat_region 15356. .15657
/note="AluX repeat: matches 1. .301 of consensus"
repeat_region 15658. .16040
/note="TIGER2 repeat: matches 863. .1266 of consensus"
repeat_region 16041. .16350
/note="AluY repeat: matches 1. .306 of consensus"
repeat_region 16351. .16417
/note="TIGER2 repeat: matches 798. .863 of consensus"
repeat_region 16445. .16538
/note="FLAM A repeat: matches 24. .117 of consensus"
repeat_region 16579. .16625
/note="TIGER2 repeat: matches 1. .47 of consensus"
repeat_region 16749. .17174
/note="MER65A repeat: matches 1. .445 of consensus"
repeat_region 17696. .17996
/note="AluX repeat: matches 1. .306 of consensus"
repeat_region 18497. .18587
/note="FLAM A repeat: matches 43. .130 of consensus"
repeat_region 18596. .18897
/note="AluX repeat: matches 1. .302 of consensus"
repeat_region 19209. .19517
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 20539. .20659
/note="MIR repeat: matches 98. .228 of consensus"
repeat_region 20912. .21038
/note="AluY repeat: matches 1. .126 of consensus"
repeat_region 21039. .21330
/note="AluX repeat: matches 1. .291 of consensus"
repeat_region 21331. .21467
/note="AluY repeat: matches 126. .266 of consensus"
repeat_region 21473. .21756
/note="AluY repeat: matches 14. .295 of consensus"
repeat_region 22240. .22524
/note="AluX repeat: matches 1. .286 of consensus"
repeat_region 23046. .23585
/note="L2 repeat: matches 2234. .2745 of consensus"
repeat_region 23679. .24214
/note="MER41A repeat: matches 1. .538 of consensus"
repeat_region 24274. .24440
/note="MIR repeat: matches 17. .191 of consensus"
repeat_region 24622. .24927
/note="AluX repeat: matches 1. .305 of consensus"
repeat_region 26536. .26843
/note="AluY repeat: matches 1. .303 of consensus"
repeat_region 27436. .27726
/note="AluY repeat: matches 13. .304 of consensus"
repeat_region 29235. .29339
/note="L2 repeat: matches 2596. .2705 of consensus"
repeat_region 29588. .29879
/note="AluX repeat: matches 1. .293 of consensus"
repeat_region 30927. .31173
/note="LTR23 repeat: matches 204. .437 of consensus"
repeat_region 31638. .31785
/note="MER5A repeat: matches 87. .189 of consensus"
repeat_region 31786. .31914
/note="FLAM C repeat: matches 1. .131 of consensus"
repeat_region 31915. .31989
/note="MER5A repeat: matches 2. .87 of consensus"
repeat_region 32211. .32511
/note="AluX repeat: matches 1. .301 of consensus"
repeat_region 33285. .33395
/note="MER81 repeat: matches 3. .114 of consensus"
repeat_region 33429. .33566
/note="FLAM C repeat: matches 1. .133 of consensus"
repeat_region 33567. .33606
/note="20 copies 2 mer tg 87% conserved"
repeat_region 34452. .34750
/note="AluX repeat: matches 1. .299 of consensus"
repeat_region 34798. .35089
/note="AluY repeat: matches 1. .292 of consensus"
repeat_region 35591. .35905
/note="AluX repeat: matches 1. .112 of consensus"
repeat_region 35917. .35978
/note="L2 repeat: matches 2622. .2690 of consensus"
repeat_region 36699. .37024
/note="AluX repeat: matches 1. .111 of consensus"
repeat_region 37215. .37349
/note="L1M5A repeat: matches 6146. .6295 of consensus"
repeat_region 37451. .37586
/note="FLAM C repeat: matches 1. .133 of consensus"
repeat_region 37718. .37742
/note="MER58B repeat: matches 73. .96 of consensus"
repeat_region 37743. .38041
/note="AluX repeat: matches 1. .301 of consensus"
repeat_region 38042. .38272
/note="MER58B repeat: matches 96. .336 of consensus"
repeat_region 38504. .38786
/note="AluX repeat: matches 6. .306 of consensus"
repeat_region 39348. .40556
/note="CpG island"
misc_feature 40556. .40556
/evidence=not experimental
misc_feature 42531. .43073
/note="match: STS: Em:G36503"
repeat_region 45639. .45763
/note="L2 repeat: matches 2617. .2750 of consensus"
repeat_region 47289. .47314
/note="13 copies 2 mer ca 100% conserved"
repeat_region 47413. .47506
Query Match 0.8%; Score 24; DB 9; Length 128960;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2191 GGAGAGAGAGTGAAGAGAGAGAA 2214
Db 90972 GGAGAGAGAGTGAAGAGAGAGAA 90995
RESULT 20
AC090561/c AC090561/c 67884 bp DNA linear HTG 03-MAR-2001
LOCUS Homo sapiens clone RP11-567N24, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC090561
ACCESSION AC090561
VERSION AC090561.1 GI:13194290
KEYWORDS HTG; HTGS; PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 67884)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-567N24
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 67884)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barra,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,

```



```

RESULT 18
AL513530/c
LOCUS       AL513530               50612 bp    DNA        linear       PRI 15-MAY-2001
DEFINITION   Human DNA sequence from clone RP11-447M22 on chromosome 6, complete
              sequence.
ACCESSION   AL513530
VERSION     AL513530.8   GI:14133035
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 50612)
AUTHORS     Direct Submission
TITLE       Submitted (15-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL     CB10 LSA, UK. E-mail enquiries: hurquay@sanger.ac.uk
            requests: clonesrequest@sanger.ac.uk
COMMENT     On May 16, 2001 this sequence version replaced gi:13751553.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em, ENBL, Ss,
            SWISSPROT, Tr, TRMBB, Wp, WORMPEP; information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C.elegans/wormpep
            This sequence
            was generated from part of bacterial clone configs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/MGP/Chr6
            RP11-447M22 is from the library RP11-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-447M22 it may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP4-708FS is at 50513 in this sequence.
            The true right end of clone RP4-682J15 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source           1..50612
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-447M22"
                     /clone_11b="RP11-11.2"
                     /note="MIR repeat: matches 88..174 of consensus"
repeat_region       1485..1570
                     /note="MIR repeat: matches 88..174 of consensus"
repeat_region       1745..1774
                     /note="15 copies 2 mer tt 86% conserved"
repeat_region       3540..3811
                     /note="ALU1E repeat: matches 265..561 of consensus"
repeat_region       3800..4122
                     /note="MLT1E repeat: matches 1..322 of consensus"
repeat_region       4169..4648
                     /note="MLT1D repeat: matches 3589..4085 of consensus"
repeat_region       5046..5161
                     /note="58 copies 2 mer aa 59% conserved"
repeat_region       6172..6350
                     /note="MERSA repeat: matches 1..189 of consensus"
repeat_region       6352..6435
                     /note="42 copies 2 mer ta 63% conserved"
repeat_region       6783..8591
                     /note="TIGGER2 repeat: matches 895..2718 of consensus"
repeat_region       8592..8882
                     /note="AluX repeat: matches 1..291 of consensus"
repeat_region       8883..9301
                     /note="TIGGER2 repeat: matches 474..895 of consensus"
repeat_region       9302..9598
                     /note="AluJb repeat: matches 1..291 of consensus"
repeat_region       9599..10107
                     /note="TIGGER2 repeat: matches 4..474 of consensus"
repeat_region       10470..10756
                     /note="AluSc repeat: matches 1..288 of consensus"
repeat_region       13276..13732
                     /note="MLT1D repeat: matches 1..505 of consensus"
repeat_region       13898..14119
                     /note="MLT1-INTERNAL repeat: matches 951..1182 of
                     consensus"
repeat_region       14554..15103
                     /note="MLT1-INTERNAL repeat: matches 36..592 of consensus"
repeat_region       15130..15428
                     /note="MLT2E repeat: matches 136..397 of consensus"
repeat_region       15401..15553
                     /note="MLT1D repeat: matches 2..154 of consensus"
repeat_region       16010..16185
                     /note="L2 repeat: matches 2191..2372 of consensus"
repeat_region       16302..16343
                     /note="L1MB8 repeat: matches 6117..6163 of consensus"
repeat_region       16353..16656
                     /note="AluX repeat: matches 1..304 of consensus"
repeat_region       16841..16908
                     /note="L2 repeat: matches 2641..2709 of consensus"
repeat_region       17903..17997
                     /note="L2 repeat: matches 2575..2680 of consensus"
repeat_region       18264..19239
                     /note="488 copies 2 mer ta 87% conserved"
repeat_region       18959
                     /note="Tandem repeat. Forced join. Gap size estimated to
                     be approximately 250bp by restriction digest data."
repeat_region       19240..19524
                     /note="AluY repeat: matches 1..285 of consensus"
repeat_region       19735..20045
                     /note="AluY repeat: matches 1..308 of consensus"
repeat_region       20344..20626
                     /note="L2 repeat: matches 2002..2294 of consensus"
repeat_region       20849..21066
                     /note="MIR repeat: matches 34..244 of consensus"
repeat_region       21538..21606
                     /note="MERSA repeat: matches 38..103 of consensus"
repeat_region       23617..23800
                     /note="MERSA repeat: matches 2..186 of consensus"
repeat_region       23953..24005
                     /note="MADB1 repeat: matches 28..80 of consensus"
repeat_region       24271..24453
                     /note="MERSA repeat: matches 2..189 of consensus"
repeat_region       26159..26323
                     /note="MIR repeat: matches 78..249 of consensus"
repeat_region       26313..26567
                     /note="L2 repeat: matches 397..654 of consensus"
repeat_region       26731..26876
                     /note="L2 repeat: matches 879..1028 of consensus"
repeat_region       27304..27596
                     /note="AluY repeat: matches 1..292 of consensus"
repeat_region       28752..29042
                     /note="L1 repeat: matches 2111..2413 of consensus"
repeat_region       29124..30699
                     /note="L1 repeat: matches 2504..3534 of consensus"
repeat_region       30174..31422
                     /note="L1M4 repeat: matches 3279..4581 of consensus"
repeat_region       31424..32086
                     /note="L1MC3 repeat: matches 7067..7739 of consensus"
repeat_region       32092..32407

```

```
TITLE
JOURNAL
COMMENT
Sequence tag and encoded human protein
Patent: JP 2001269182-A 29842 02-OCT-2001;
GENSET
OS Homo sapiens (human)
PN JP 2001269182-A/29842
PD 02-OCT-2001 JP 2000118773
PR 24-FEB-2000 JP 2000118773
PI JEAN BAPTISTE DUMAS MITLINE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
PI JORDAN
PC C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21, PC
C12N5/10,
PC C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00, PC
G06F15/40
CC
FH Key Location/Qualifiers.
FEATURES
source
1..246
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 70 a 63 c 71 g 40 t 2 others
ORIGIN
Query Match 4.6%; Score 131; DB 6; Length 246;
Best Local Similarity 100.0%; Pred. No. 1.5e-59;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2722 GACCAGGTCCTCCCTGGCAAGATGGAGACCATGAAAACCTGGAAATCCAAAGGCAACCA 2751
DB 112 GACCAGGTCCTCCCTGGCAAGATGGAGACCATGAAAACCTGGAAATCCAAAGGCAACCA 171

QY 2782 GCCCCCCCAGGATCTGGACCCATCACTATGTTTGTAGTGAATGGCAGAGAGATCCG 2841
DB 172 GCCCCCCCAGGATCTGGACCCATCACTATGTTTGTAGTGAATGGCAGAGAGATCCG 231

QY 2842 TTCAGAAAAGG 2852
DB 232 TTCAGAAAAGG 242

RESULT 16
AX721212 1101 bp mRNA linear PAT 07-MAY-2003
LOCUS
DEFINITION
Sequence 172 from Patent WO0220754.
ACCESSION
AX721212
VERSION
AX721212.1 GI:30422048
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Lincoln,S.E., Altus,C.M., Dufour,G.E., Chalup,M.S., Hillman,J.L.,
Jones,A., Yu,J.Y., Wright,R.J., Gietzen,D., Liu, Tommy P., Yap,P.,
Dahl,C.R., Moniyama,M., Bradley,D., Rohatgi,S., Harris,B.,
Roseberry,A.M., Gerstin,E.H., Petalta,C.H., David,M., Panzer,S.,
Flores,V., Dafo,A., Marwaha,R., Chen,A., Chang,S.C. and Imman,R.R.
Molecules for diagnostics and therapeutics
Patent: WO 0220754-A 172 14-MAR-2002;
INCYTE
IncYTE Genomics, Inc. (US)
FEATURES
source
1..1101
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
BASE COUNT 334 a 191 c 203 g 373 t
ORIGIN
Query Match 2.2%; Score 63; DB 6; Length 1101;
Best Local Similarity 100.0%; Pred. No. 3.2e-22;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1586 AGGGTGAATGGTGCCAAAGGAGACAAAGATCACTCGTTTATGCGCAAAAGGG 1643
DB 2625 AGGGTGAATGGTGCCAAAGGAGACAAAGATCACTCGTTTATGCGCAAAAGGG 2682
```

```
QY 2195 AGAGAGGTGAAAAGGGAGAACTCGGTGTCGAGTGCCATTGGATCAAAAGGAGATCTG 2254
|||||
DB 140 AGAGAGGTGAAAAGGGAGAACTCGGTGTCGAGTGCCATTGGATCAAAAGGAGATCTG 199
|||||

QY 2255 GGG 2257
||||
DB 200 GGG 202

RESULT 17
AL450431 25483 bp DNA linear PRI 30-MAY-2002
LOCUS
DEFINITION
Human DNA sequence from clone RP11-323114 on chromosome 6, complete
sequence.
ACCESSION
AL450431 AL39091
VERSION
AL450431.1 GI:11493128
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 25483)
Direct Submission
Submitted (28-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP11-323114 is from
the library RPCI-11.2 constructed by the group of Pieter de Jong.
For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.
FEATURES
source
1..25483
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-323114"
/clone_lib="RPCI-11.2"
BASE COUNT 7053 a 4365 c 4610 g 9455 t
ORIGIN
Query Match 2.0%; Score 58; DB 9; Length 25483;
Best Local Similarity 100.0%; Pred. No. 1.4e-19;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1586 AGGGTGAATGGTGCCAAAGGAGACAAAGATCACTCGTTTATGCGCAAAAGGG 1643
DB 2625 AGGGTGAATGGTGCCAAAGGAGACAAAGATCACTCGTTTATGCGCAAAAGGG 2682
```

Db 205 ATCTTAGTGGCTTAATTTGTTGGCCAGAAAATTTGAATAGTGAAGTGGCTT 264
 QY 181 GTCAATATCACAATAACTTTGACATAGGCCCAAGCTTATTCAGTTGAGTGGTTCAA 240
 Db 265 GTCATATCACAATAACTTTGACATAGGCCCAAGCTTATTCAGTTGAGTGGTTCAA 324
 QY 241 TATAGTGAATACCTGTGCTGGAGATTCCTCTCGAAGCTATGATTCAGGAGAATTTG 300
 Db 325 TATAGTGAATACCTGTGCTGGAGATTCCTCTCGAAGCTATGATTCAGGAGAATTTG 384
 QY 301 ACCGACAGCTGGATCCATCTACTTAGGAGGAAACACAAAGACAGGGAG 354
 Db 385 ACCGACAGCTGGATCCATCTACTTAGGAGGAAACACAAAGACAGGGAG 438

RESULT 13
 AF438327_3
 WPCOMMENT
 Sequence split into 4 fragments LOCUS AF438327 Accession AF438327
 Fragment Name Begin End
 AF438327_0 1 110000
 AF438327_1 100001 210000
 AF438327_2 200001 310000
 AF438327_3 300001 381696

Continuation (4 of 4) of AF438327 from base 300001 (AF438327 Homo sapiens alpha 1 type X

Query Match 7.1%; Score 204; DB 9; Length 81696;
 Best Local Similarity 100.0%; Pred. No. 6.9e-100;
 Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2397 AGCCAGCTACCACTCTTACTTCAGAGTGAAGAATTAGAAATTTGATCATTTGCTGTC 2456
 Db 47106 AGCCAGCTACCACTCTTACTTCAGAGTGAAGAATTAGAAATTTGATCATTTGCTGTC 47165
 QY 2457 CCACATGGCTCCCGGGTATTCCTGGCCACCTGTCGATAGGCCAGAGGTCGCCAG 2516
 Db 47166 CCACATGGCTCCCGGGTATTCCTGGCCACCTGTCGATAGGCCAGAGGTCGCCAG 47225
 QY 2517 AGGATTACCTGTTTGCAGGAGAGATGGTGTCTCTGGATTAGTGGGTGTCCTGGAGC 2576
 Db 47226 AGGATTACCTGTTTGCAGGAGAGATGGTGTCTCTGGATTAGTGGGTGTCCTGGAGC 47285
 QY 2577 TCCAGGTGTCAGAGATTAAAGG 2600
 Db 47286 TCCAGGTGTCAGAGATTAAAGG 47309

RESULT 14
 HS682J15/c
 LOCUS
 DEFINITION Human DNA sequence from clone RP4-682J15 on chromosome 6p11.2-12.3, complete sequence.
 ACCESSION AL034452
 VERSION AL034452.8 GI:4678496
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Smith.S.
 Direct Submission
 Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Apr 24, 1999 this sequence version replaced gi:4586169.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Sanger code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one X13 subclone; and the rare assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/Celegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> RP4-682J15 is from the library RPCI-4 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pCYPAC2

FEATURES
 source
 1..118431
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosomes="6"
 /map="p11.2-12.3"
 /clone="RP4-682J15"
 /scler_lib="RPCI-4"

BASE COUNT 38100 a 21308 c 21316 g 37707 t
 ORIGIN
 Query Match 7.1%; Score 204; DB 9; Length 118431;
 Best Local Similarity 100.0%; Pred. No. 6.6e-100;
 Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2397 AGCCAGCTACCACTCTTACTTCAGAGTGAAGAATTAGAAATTTGATCATTTGCTGTC 2456
 Db 99561 AGCCAGCTACCACTCTTACTTCAGAGTGAAGAATTAGAAATTTGATCATTTGCTGTC 99002
 QY 2457 CCACATGGCTCCCGGGTATTCCTGGCCACCTGTCGATAGGCCAGAGGTCGCCAG 2516
 Db 99501 CCACATGGCTCCCGGGTATTCCTGGCCACCTGTCGATAGGCCAGAGGTCGCCAG 98942
 QY 2517 AGGATTACCTGTTTGCAGGAGAGATGGTGTCTCTGGATTAGTGGGTGTCCTGGAGC 2576
 Db 98941 AGGATTACCTGTTTGCAGGAGAGATGGTGTCTCTGGATTAGTGGGTGTCCTGGAGC 98882
 QY 2577 TCCAGGTGTCAGAGATTAAAGG 2600
 Db 98881 TCCAGGTGTCAGAGATTAAAGG 98858

RESULT 15
 BD053596
 LOCUS
 DEFINITION Sequence tag and encoded human protein.
 ACCESSION BD053596
 VERSION BD053596.1 GI:22599202
 KEYWORDS GP 2001269182-A/29842.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Edwards.J.B.D.M., Duclair,E. and Jordan,J.Y.
 REFERENCE 1 (bases 1 to 246)
 AUTHORS

```

QY 209 GCGCGAAGTTTATTCAAGTTGGAGTGGTTCAATATATAGTGACTACCTGTGCTGGAGATTC 268
Db 50007 GCGCGAAGTTTATTCAAGTTGGAGTGGTTCAATATATAGTGACTACCTGTGCTGGAGATTC 49948
QY 269 CTCTCGGAAGCTATGATTTCAGAGAAACATTTTCAGCGCAGCAGTGGAAATCCATCTACT 328
Db 49947 CTCTCGGAAGCTATGATTTCAGAGAAACATTTTCAGCGCAGCAGTGGAAATCCATCTACT 49888
QY 329 TAGGAGAAACACAAAGACAGGAGGAGCCATCAGTTTCGGCTCGCATTTTTCGCA 388
Db 49887 TAGGAGAAACACAAAGACAGGAGGAGCCATCAGTTTCGGCTCGCATTTTTCGCA 49828
QY 389 AGTCCTCACCATTCTGACTTAAGATAGCAGTGTCTTACGATGCGAAATCCCAAGATG 448
Db 49827 AGTCCTCACCATTCTGACTTAAGATAGCAGTGTCTTACGATGCGAAATCCCAAGATG 49768
QY 449 ACCTCAAGGATGAGCTCAAGCAGCAGAGATAGTAGAATAACATTTATTTGCTATTGGTG 508
Db 49767 ACCTCAAGGATGAGCTCAAGCAGCAGAGATAGTAGAATAACATTTATTTGCTATTGGTG 49708
QY 509 TTGGTTTCAGAAACAGAGATGCCAACTTAGAGCTATTGCCAAGAGCTTCGCTACTT 568
Db 49707 TTGGTTTCAGAAACAGAGATGCCAACTTAGAGCTATTGCCAAGAGCTTCGCTACTT 49648
QY 569 ATGTGTTTTATGTGGAGACTATATTGCAATATCCAAATAAGGGAAGTGTATGAAGCAGA 628
Db 49647 ATGTGTTTTATGTGGAGACTATATTGCAATATCCAAATAAGGGAAGTGTATGAAGCAGA 49588
QY 629 AACTTTGTGAAG 640
Db 49587 AACTTTGTGAAG 49576

RESULT 11
AX721209 1414 bp mRNA linear PAT 07-MAY-2003
LOCUS
DEFINITION Sequence 169 from Patent WO0220754.
ACCESSION AX721209
VERSION AX721209.1 GI:30422045
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens

REFERENCE
AUTHORS Lincoln,S.E., Altus,C.M., Dufour,G.E., Chalup,M.S., Hillman,J.L., Jones,A., Yu,J.Y., Wright,R.J., Gietzen,D., Liu, Toomy,F., Yap,P., Dahl,C.R., Momiyama,M., Bradley,D., Rohatgi,S., Harris,B., Roseberry,A.M., Gerstin,E.H., Peralta,C.H., David,M., Panzer,S., Flores,V., Daffo,A., Marwaha,R., Chen,A., Chang,S.C. and Inman,R.R.
TITLE Molecules for diagnostics and therapeutics
JOURNAL Patent: WO 0220754-A 169 14-MAR-2002;
INCYTE Genomics, Inc. (US)
FEATURES
source
1. .1414
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon.9606"
/notes="Incyte ID No: LG:407788.2:2000SEP08"
BASE COUNT 425 a 267 c 295 g 427 t
ORIGIN
Query Match 13.3%; Score 382; DB 6; Length 1414;
Best Local Similarity 100.0%; Pred. No. 1.2e-197;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2397 AGCCAGCTACCACTCTTACTTCAGAGTGAAGAAATTAGAAATTTGTGATCATTCGCTGTC 2456
Db 959 AGCCAGCTACCACTCTTACTTCAGAGTGAAGAAATTAGAAATTTGTGATCATTCGCTGTC 1018
QY 2457 CCAACATGGCTCCCGGGTATTCTCTGGCCACCTGGTCGATAGGCCAGAGGTCGCCAG 2516
Db 959 CCAACATGGCTCCCGGGTATTCTCTGGCCACCTGGTCGATAGGCCAGAGGTCGCCAG 1018

Db 1019 CCAACATGGCTCCCGGGTATTCTCTGGCCACCTGGTCGATAGGCCAGAGGTCGCCAG 1079
QY 2517 AGGATTACCTGGTTTTCAGAAAGAGATGGTTTCTTGGATTAGTGGTGTCTCTCGAAG 2576
Db 1078 AGGATTACCTGGTTTTCAGAAAGAGATGGTTTCTTGGATTAGTGGTGTCTCTCGAAG 1138
QY 2577 TCCAGGTGTCAGAGGATTAAAGGCCCTACAGGAGAAATCGGGAAGAGGAGCCCAAGS 2636
Db 1139 TCCAGGTGTCAGAGGATTAAAGGCCCTACAGGAGAAATCGGGAAGAGGAGCCCAAGS 1198
QY 2637 GTTTCGGTATCTCTGGAAACAAGGTCTCTCTGGTGTCTCTGGTGTCTCTGGTGTCTCTGG 2696
Db 1199 GTTTCGGTATCTCTGGAAACAAGGTCTCTCTGGTGTCTCTGGTGTCTCTGGTGTCTCTGG 1258
QY 2697 AATAAGCAAGAGGTCTCTCAGGAGACCCAGGTCTCTCTGGTGTCTCTGGTGTCTCTGGTGTCTCTGG 2756
Db 1259 AATAAGCAAGAGGTCTCTCAGGAGACCCAGGTCTCTCTGGTGTCTCTGGTGTCTCTGGTGTCTCTGG 1318
QY 2757 AAAACCTGGAATCCCAAGGGCAA 2778
Db 1319 AAAACCTGGAATCCCAAGGGCAA 1340

RESULT 12
BD058397
LOCUS
DEFINITION Secreted expressed sequence tags (seSTs).
ACCESSION BD058397
VERSION BD058397.1 GI:22604003
KEYWORDS JP 2001519666-A/252.
SOURCE
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 1 to 447)
JACOBS,K., MCCOY,J.M., LAVALLIE,E.R., RACIE,L.A., VERBERG,D., TREACY,M., SPAULDING,V. and AGOSTINO,M.J.
Secreted expressed sequence tags (seSTs)
Patent: JP 2001519666-A 252 23-OCT-2001;
GENETICS INSTITUTE INC
COMMENT
PN JP 2001519666-A/252
EP 23-OCT-2001
EP 10-APR-1998 JP 1998543068
PR 10-APR-1997 US 08/835513
PI KENNETH JACOBS,JOHN M MCCOY,EDWARD R LAVALLIE,LISA A RACIE,PI DAVID VERBERG,
PI MAURICE TREACY,VIRKI SPAULDING,MICHAEL J AGOSTINO PC
C12N15/12,C12N5/10,C07K14/47,C12Q1/68,A61K38/17 CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
1. .447
/organism="Zea mays"
/mol_type="genomic DNA"
/db_xref="taxon:4577"
BASE COUNT 141 a 80 c 101 g 125 t
ORIGIN
Query Match 8.8%; Score 252; DB 6; Length 447;
Best Local Similarity 99.4%; Pred. No. 4.1e-126;
Matches 352; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTCACTATATTACATTTCTCTGCATGTTTTGGTGTCTCTCAGATTCTGTG 60
Db 85 ATGGCTCACTATATTACATTTCTCTGCATGTTTTGGTGTCTCTCAGATTCTGTG :44
QY 61 TTAGCTGAAGATGGGAAGTAAAGATCAAGTTGCTGTACTGCTCCGACAGATTAGTTTC 120
Db 145 TTAGCTGAAGATGGGAAGTAAAGATCAAGTTGCTGTACTGCTCCGACAGATTAGTTTC 204
QY 121 ATCTTAGATGGCTCTTAGTGTGGCCGCAAAACTTTGAATAGTGAAGAAAGAGGCTT 180

```

Db	915	AAAAAGTTTAAGAAAGATACAGCTTTTCCAAAAAAGATATAAGGATATGAAGTAACA	974
Qy	781	TCAAAAGTTGATTATATCAAGACTCAAGCAATGTTTCCCAAGAGGCTTCCTCCCATCA	840
Db	975	TCAAAAGTTGATTATATCAAGACTCAAGCAATGTTTCCCAAGAGGCTTCCTCCCATCA	1034
Qy	841	TATGATTTGTGCTACTCAAGATTAAAGTCAAGAAATTTGGATTTATGAGAATA	900
Db	1035	TATGATTTGTGCTACTCAAGATTAAAGTCAAGAAATTTGGATTTATGAGAATA	1094
Qy	901	TTAACTATTGATGGAAGGCCA	921
Db	1095	TTAACTATTGATGGAAGGCCA	1115
RESULT 9			
AF438327.2			
WPCOMMENT			
Sequence split into 4 fragments LOCUS AF438327 Accession AF438327			
Fragment Name Begin End			
AF438327_0 1 110000			
AF438327_1 100001 210000			
AF438327_2 200001 310000			
AF438327_3 300001 381696			
Continuation (3 of 4) of AF438327 from base 200001 (AF438327 Homo sapiens alpha 1 type X			
Query Match 19.3%; Score 552; DB 9; Length 110000;			
Best Local Similarity 100.0%; Pred. No. 3e-291;			
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	89	GTTCGTACTGCTCCGACAGATTAGTTTTCATCTTAGTGCGCTCTTATAGTGTGCC	148
Db	27182	GTTCGTACTGCTCCGACAGATTAGTTTTCATCTTAGTGCGCTCTTATAGTGTGCC	27241
Qy	149	CAGAAACTTTGAATAGTCAAAAGTGGCTTGTCAATATACAAAAACCTTGACATAG	208
Db	27242	CAGAAACTTTGAATAGTCAAAAGTGGCTTGTCAATATACAAAAACCTTGACATAG	27301
Qy	209	GGCCGAAGTTTATCAAGTTGGAGTGTTCATATAGTGACTACCTGTGCTGGAGATTC	268
Db	27302	GGCCGAAGTTTATCAAGTTGGAGTGTTCATATAGTGACTACCTGTGCTGGAGATTC	27361
Qy	269	CTCTCGAAGTATGATTCAGGAGACATTTGACGCGACAGTGGATCCATCTACT	328
Db	27362	CTCTCGAAGTATGATTCAGGAGACATTTGACGCGACAGTGGATCCATCTACT	27421
Qy	329	TAGAGGAAACACAAAGACAGGAGGCCATCCAGTTTTCGCTCGATTACCTTTTGCCA	388
Db	27422	TAGAGGAAACACAAAGACAGGAGGCCATCCAGTTTTCGCTCGATTACCTTTTGCCA	27481
Qy	389	AGTCCTCAGATTCTGACTTAAGATAGCAGTGTACTTACGATGCGAAATCCGAGATG	448
Db	27482	AGTCCTCAGATTCTGACTTAAGATAGCAGTGTACTTACGATGCGAAATCCGAGATG	27541
Qy	449	ACGTCAAGGATGAGCTCAAGCAGCAGAGATAGTAAATATATTTGCTATTTGGTG	508
Db	27542	ACGTCAAGGATGAGCTCAAGCAGCAGAGATAGTAAATATATTTGCTATTTGGTG	27601
Qy	509	TTGGTTCAGAAACAGAGATGCCAATTTAGACTTATTCGCAACAGCTTCGCTACTT	568
Db	27602	TTGGTTCAGAAACAGAGATGCCAATTTAGACTTATTCGCAACAGCTTCGCTACTT	27661
Qy	569	ATGTGTTTTATGTGGAAGACTATATTGCAATATCCAAATTAAGGGAAGTATGAAGCAGA	628
Db	27662	ATGTGTTTTATGTGGAAGACTATATTGCAATATCCAAATTAAGGGAAGTATGAAGCAGA	27721
Qy	629	AACTTTGTGAAG	640
Db	27722	AACTTTGTGAAG	27733
RESULT 10			
HS708F5/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
<p>HS708F5 126501 bp DNA linear PRI 04-MAR-2003</p> <p>Human DNA sequence from clone RP4-708F5 on chromosome 6p11.2-12.3, complete sequence.</p> <p>AL031782</p> <p>AL031782.1 GI:4164360</p> <p>HTG.</p> <p>Homo sapiens (human)</p> <p>Homo sapiens</p> <p>Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.</p> <p>1. (bases 1 to 126501)</p> <p>Griffiths, C.</p> <p>Direct Submission</p> <p>Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk</p> <p>Humquery@sanger.ac.uk</p> <p>On Jan 19, 1999 this sequence version replaced gi:3947543.</p> <p>----- Genome Center</p> <p>Center: Wellcome Trust Sanger Institute</p> <p>Center code: SC</p> <p>Web site: http://www.sanger.ac.uk</p> <p>Contact: humquery@sanger.ac.uk</p> <p>-----</p> <p>During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.</p> <p>This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.</p> <p>The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep</p> <p>This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6</p> <p>RP4-708F5 is from the library RPI-4 constructed by the group of Pieter de Jong. For further details see http://www.ctcri.org/bacpac/home.htm</p> <p>VECTOR: pCVPAC2.</p>			
FEATURES			
source			
<p>1..126501</p> <p>Location/Qualifiers</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:9606"</p> <p>/chromosome="6"</p> <p>/map="p11.2-12.3"</p> <p>/clone="RP4-708F5"</p> <p>/clone_lib="RPI-4"</p>			
BASE COUNT 44177 a 23609 c 21302 g 36813 t			
ORIGIN			
<p>Query Match. 19.3%; Score 552; DB 9; Length 126501;</p> <p>Best Local Similarity 100.0%; Pred. No. 2.9e-291;</p> <p>Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>			
Qy	89	GTTCGTACTGCTCCGACAGATTAGTTTTCATCTTAGTGCGCTCTTATAGTGTGCC	148
Db	50127	GTTCGTACTGCTCCGACAGATTAGTTTTCATCTTAGTGCGCTCTTATAGTGTGCC	50068
Qy	149	CAGAAACTTTGAATAGTCAAAAGTGGCTTGTCAATATACAAAAACCTTGACATAG	208
Db	50067	CAGAAACTTTGAATAGTCAAAAGTGGCTTGTCAATATACAAAAACCTTGACATAG	50008

Db	349	GGGATCCCTGGGGCTCTGGGCTCAAGGGAGAACCCAGGAGCAACGGGTCCCGCAGGAGAA	408
Qy	2044	CCAGGATACATGGGTTTACCCGGGATTTCAAGGAAAAAGGGGCAAAAGGAAATCAAGGT	2103
Db	409	CCAGGATACATGGGTTTACCCGGGATTTCAAGGAAAAAGGGGCAAAAGGAAATCAAGGT	468
Qy	2104	GAAGAGGTTATTCAGGTCGAAAGGGGAGAAATGGAAGACAGGGGAATTCAGGGCAACAG	2163
Db	469	GAAGAGGTTATTCAGGTCGAAAGGGGAGAAATGGAAGACAGGGGAATTCAGGGCAACAG	528
Qy	2164	GGATTCAGAGCCATCATGGTCAAAAGGAGAGAGAGGTGAAGAGGAGAAACCTGGGTGC	2223
Db	529	GGATTCAGAGCCATCATGGTCAAAAG--GAGAGAGGTGAAGAGGAGAACCTGGGTGC	586
Qy	2224	CGAGGTGCCATTCAGATCAAAAGGAGAAATCTGGGGTGGATGGCTTGATGGGCGCCGAGGT	2283
Db	587	CGAGGTGCCATTCAGATCAAAAGGAGAAATCTGGGGTGGATGGCTTGATGGGCGCCGAGGT	646
Qy	2284	CCTAAGGGGCAACCTCGGATCCAGGTCTCAGGGACCCCGCAGGTTCGGATGGGAAGGCC	2343
Db	647	CCTAAGGGGCAACCTCGGATCCAGGTCTCAGGGACCCCGCAGGTTCGGATGGGAAGGCC	706
Qy	2344	GGAAGAGGTTTTCAGAAATTTATTCAGCAAGTTTGCACAGATGTAAATAGAGCCGAG	2403
Db	707	GGAAGAGGTTTTCAGAAATTTATTCAGCAAGTTTGCACAGATGTAAATAGAGCCGAG	766
Qy	2404	CTACCACTTACTTCAGAGTGAAGAAATTCAGAAATTTGATCATCTGTGCCCAAGAT	2463
Db	767	CTACCACTTACTTCAGAGTGAAGAAATTCAGAAATTTGATCATCTGTGCCCAAGAT	826
Qy	2464	GGCTCCCGGGTATTCCTGGGCCACCTGGTCCGATAGGCCCGCAGAGGGTCCAGAGGATTA	2523
Db	827	GGCTCCCGGGTATTCCTGGGCCACCTGGTCCGATAGGCCCGCAGAGGGTCCAGAGGATTA	886
Qy	2524	CCTGGTTGGCAGAGAGATGGTGTCTCGATAGTGGTGTCTCGTGGAGTCCAGGT	2583
Db	887	CCTGGTTGGCAGAGAGATGGTGTCTCGATAGTGGTGTCTCGTGGAGTCCAGGT	946
Qy	2584	GTGAGAGGATTAAGAGCCCTTACCAGGAGAAATGGGGAAGAGGAGGCAAGGGTTGGG	2643
Db	947	GTGAGAGGATTAAGAGCCCTTACCAGGAGAAATGGGGAAGAGGAGGCAAGGGTTGGG	1006
Qy	2644	TATCTGGAGAACAGGTCTCTCTGGTCCCGCAGGTCCAGAGGGCCCTCTCGGAATAGC	2703
Db	1007	TATCTGGAGAACAGGTCTCTCTGGTCCCGCAGGTCCAGAGGGCCCTCTCGGAATAGC	1066
Qy	2704	AAAGAGGTCCTCCAGGAGACCCAGGTCTCTCTGGTCCCGCAGGTCCAGAGGCAAGCT	2763
Db	1067	AAAGAGGTCCTCCAGGAGACCCAGGTCTCTCTGGTCCCGCAGGTCCAGAGGCAAGCT	1126
Qy	2764	GGATCCAGGGGCAACAGGCCCCCAGGATCTCGAGCCGATCACTATGTTAGTGA	2823
Db	1127	GGATCCAGGGGCAACAGGCCCCCAGGATCTCGAGCCGATCACTATGTTAGTGA	1186
Qy	2824	ATTGCCAGAGAGATCGTTTCAGAAAAGGACCAAACTATTAG	2865
Db	1187	ATTGCCAGAGAGATCGTTTCAGAAAAGGACCAAACTATTAG	1228
RESULT 8			
AX552238		1115 bp	DNA
LOCUS			linear
DEFINITION	Sequence 147 from Patent WO0162927.		PAT 27-NOV-2002
ACCESSION	AX552238		
VERSION	AX552238.1		GI:25896476
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1		
AUTHORS	Banville,S.C., Greenawalt,L.B., Lincoln,S.E., Stockdreher,T.K.,		

Amshay,S., Chang,S.C., Chen,W., D'Sa,S.A., Dam,T.C., Liu,T.F.,			
Rosen,B.H., Russo,F.D., Spiro,P.A., Bradley,D.L., Chen,A.,			
Cohen,H.J., Daffo,A., Daniels,S.E., Dufour,G.E., Flores,V.,			
Fong,W.T., Hodgson,D.M., Jackson,S., Jones,A.L., Panzer,S.,			
Roseberry,A.M., Shah,P., Wright,R.J., Yap,P.E., Yu,J.Y.,			
Bratcher,S.R., Chalup,M.S., Dahl,C.R. and Hillman,J.L.			
Polypeptides and corresponding polynucleotides for diagnostics and			
therapeutics			
Patent: WO 0162927-A 147 30-AUG-2001;			
Incyte Genomics, Inc. (US)			
Location/Qualifiers			
1..1115			
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
/note="Incyte ID No: LG:120744.1:2000MAY19"			
BASE COUNT	344 a	215 c	259 g
ORIGIN			
Query Match	26.1%	Score 749;	DB 6; Length 1115;
Best Local Similarity	99.8%	Pred. No. 0;	
Matches	919;	Conservative 0;	Mismatches 1; Gaps 1;
Qy	1	ATGGCTCACTATATACATTTCTCTGCATGGTTTGGTGTGCTTCTTCAGAAATCTGTG	60
Db	196	ATGGCTCACTATATACATTTCTCTGCATGGTTTGGTGTGCTTCTTCAGAAATCTGTG	255
Qy	61	TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGCTGCTCCGACAGATTTAGTTTC	120
Db	256	TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGCTGCTCCGACAGATTTAGTTTC	315
Qy	126	ATCTTAGATGGCTCTTATAGTGTGGCCACAGAAACTTGAATAGTGAAGTGGCT	180
Db	316	ATCTTAGATGGCTCTTATAGTGTGGCCACAGAAACTTGAATAGTGAAGTGGCT	375
Qy	181	GTCAATATCAACAAAACCTTTGACATAGGCGGAGGTTTATCAAGTGGAGTGGTCAA	240
Db	376	GTCAATATCAACAAAACCTTTGACATAGGCGGAGGTTTATCAAGTGGAGTGGTCAA	435
Qy	241	TATAGTGACTACCTCTGCTCGAGATTCCTCTCGAAAGCTATGATTCAGGAGAACATTTG	300
Db	436	TATAGTGACTACCTCTGCTCGAGATTCCTCTCGAAAGCTATGATTCAGGAGAACATTTG	495
Qy	301	ACGGCAGCAGTGGAAATCCATCTCTTACAGAGAAACACAAAGACAGGAGGCGCATC	360
Db	496	ACGGCAGCAGTGGAAATCCATCTCTTACAGAGAAACACAAAGACAGGAGGCGCATC	555
Qy	361	CAGTTTGGCTCGATTTACCTTTTGGCCAAAGTCTCAGGATTTCTGACTAAGATAGCAGTG	420
Db	556	CAGTTTGGCTCGATTTACCTTTTGGCCAAAGTCTCAGGATTTCTGACTAAGATAGCAGTG	675
Qy	421	GTACTTAGGATGGCAAAATCCAGATGACGTCAAGATGCGAGTCCAGCAGCAAGAT	480
Db	616	GTACTTAGGATGGCAAAATCCAGATGACGTCAAGATGCGAGTCCAGCAGCAAGAT	675
Qy	481	AGTAAGATAACATTTATTTGCTATTGGTGTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTT	540
Db	676	AGTAAGATAACATTTATTTGCTATTGGTGTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTT	735
Qy	541	GCTATTGCCAACAAGCCCTTCTGCTTACTTATGTTTATGTTTATGTTTATGTTTATGTTT	600
Db	736	GCTATTGCCAACAAGCCCTTCTGCTTACTTATGTTTATGTTTATGTTTATGTTTATGTTT	795
Qy	601	TCCAAAATTAAGGGAAGTGAAGCAGAAACCTTTGTGAAGAATCTGCTCTGTCACCAACGA	660
Db	796	TCCAAAATTAAGGGAAGTGAAGCAGAAACCTTTGTGAAGATCTGCTCTGTCACCAACGA	855
Qy	661	ATTCCAGTGGCAGCTTCGTGATGAAGGGGATTTGATATTTCTTTTGTAGTTTGTAAAT	720
Db	856	ATTCCAGTGGCAGCTTCGTGATGAAGGGGATTTGATATTTCTTTTGTAGTTTGTAAAT	914
Qy	721	AAAAAGGTTAAGAAAGATACAGCTTTCACCAAAAAAGATTAAGATATGAATGAACA	780

```
source 1..1169
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/notes="Incyte ID No: LG:120744.1:2000SEP08"
BASE COUNT 364 a 224 c 267 g 314 t
ORIGIN
Query Match 32.2%; Score 923; DB 6; Length 1169;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 973; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGGCTCACTATATACATTTCTCTGCATGGTTTGGTGTCTCTTTCAGAAATTCGTGG 60
Db 196 ATGGCTCACTATATACATTTCTCTGCATGGTTTGGTGTCTCTTTCAGAAATTCGTGG 255
Qy 61 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGTCGTACTGCTCCGACAGATTTAGTTTC 120
Db 256 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGTCGTACTGCTCCGACAGATTTAGTTTC 315
Qy 121 ATCTAGATGGCTCTTATAGTTTGGCCAGAAAACCTTTGAATATAGTGAAGTGGCTT 180
Db 316 ATCTAGATGGCTCTTATAGTTTGGCCAGAAAACCTTTGAATATAGTGAAGTGGCTT 375
Qy 181 GTCAATATCAAAAAAATTTGACATAGGCGCGAAGTTTATTCAAGTTGGAGTGGTTCAA 240
Db 376 GTCAATATCAAAAAAATTTGACATAGGCGCGAAGTTTATTCAAGTTGGAGTGGTTCAA 435
Qy 241 TATAGTGAATACCTGCTGGAGATTCCTCTCGGAAGCTATGATTCAGAGAAATTTG 300
Db 436 TATAGTGAATACCTGCTGGAGATTCCTCTCGGAAGCTATGATTCAGAGAAATTTG 495
Qy 301 ACGCAGCAGTGGATCCATCTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 360
Db 496 ACGCAGCAGTGGATCCATCTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 555
Qy 361 CAGTTGGCTCGATTACCTTTTGGCCAGTCTCTCAGATTTCTGACTAAGTAGAGTG 420
Db 556 CAGTTGGCTCGATTACCTTTTGGCCAGTCTCTCAGATTTCTGACTAAGTAGAGTG 615
Qy 421 GTACTTACGATGCAATCCAGATGAGTCAAGATGACGTCAGATGACGTCACGACGACAGAT 480
Db 616 GTACTTACGATGCAATCCAGATGAGTCAAGATGACGTCAGATGACGTCACGACGACAGAT 675
Qy 481 AGTAAGATTAACATTTATGCTATTTGGTGTGGTTCAGAAACAGAGATCCGAACTTGA 540
Db 676 AGTAAGATTAACATTTATGCTATTTGGTGTGGTTCAGAAACAGAGATCCGAACTTGA 735
Qy 541 GCTATTGCAACAGCCCTTCTGCTACTATTATGTTTATGTTGGAAGACTATATGCAATA 600
Db 736 GCTATTGCAACAGCCCTTCTGCTACTATTATGTTTATGTTGGAAGACTATATGCAATA 795
Qy 601 TCCAAAATTAAGGAAGTGAAGACAGAAACTTTGTSAAGATCTCTGTCTCCACACGA 660
Db 796 TCCAAAATTAAGGAAGTGAAGACAGAAACTTTGTSAAGATCTCTGTCTCCACACGA 855
Qy 661 ATCCAGTGGAGCTCGTGAAGGAGATTTGATATCTTTTAGTCTTGAATCTTAT 720
Db 856 ATCCAGTGGAGCTCGTGAAGGAGATTTGATATCTTTTAGTCTTGAATCTTAT 915
Qy 721 AAAAAGGTTAAGAAAGAAATACAGCTTTTCCAAAAAAGATAAAGGATATGAAGTAACA 780
Db 916 AAAAAGGTTAAGAAAGAAATACAGCTTTTCCAAAAAAGATAAAGGATATGAAGTAACA 975
Qy 781 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTCCAGAGGCTTCTCTCCATCA 840
Db 976 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTCCAGAGGCTTCTCTCCATCA 1035
Qy 841 TATGATTTTGTCTTACTCAAGATTTTAAAGTCAAGAAATTTGGGATTTATGGGAATA 900
Db 1036 TATGATTTTGTCTTACTCAAGATTTTAAAGTCAAGAAATTTGGGATTTATGGGAATA 1095
Qy 901 TTAACATATTGATGGAGGCCCAAAATAGCAGTTACCTTAAATGTTGTGGACAAAATCTTTA 960
```

```
Db 1096 TTAACATATTGATGGAGGCCCAAAATAGCAGTTACCTTAAATGTTGTGGACAAAATCTTTA 1155
Qy 961 TTATTATACACAAAC 974
Db 1156 TTATTATACACAAAC 1169

RESULT 7
AK096444 2295 bp tRNA linear PRI 15-JUL-2002
LOCUS Homo sapiens cDNA FLJ33525 fis, clone NTONG2607034.
DEFINITION AK096444
ACCESSION AK096444.1 GI:21755943
VERSION 1
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
Arita,M., Musashino,K., Yuki,H., Hara,H., Sugiyama,T., Irie,R.,
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Sono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,
Suzuki,Y., Sugano,S., Nagahara,K., Masuko,Y., Nagai,K. and
Isogai,T.
TITLE NECO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2295)
AUTHORS Isogai,T. and Yamamoto,J.
DIRECT SUBMISSION
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
E-mail:genomics@hri.co.jp, tel:81-438-52-3975, Fax:81-438-52-3986;
NECO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology (RAB); cDNA library
construction; Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing; RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing; HRI and
RAB; annotation: HRI and RAB.
FEATURES
Location/Qualifiers
1..2295
/organism="Homo sapiens"
/mol_type="tRNA"
/db_xref="taxon:9606"
/clone="NTONG2007034"
/tissue_type="tongue"
/clone_lib="NTONG2"
/notes="cloning vector: pME18SFL3"
BASE COUNT 683 a 427 c 565 g 620 t
ORIGIN
Query Match 30.7%; Score 880; DB 9; Length 2295;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1060; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
Qy 1804 GGAATCCCAGGATTTCTCGAAACCGAGATTAATGGGCCAAAAGGAGAAATGGGCT 1863
Db 169 GGAATCCCAGGATTTCTCGAAACCGAGATTAATGGGCCAAAAGGAGAAATGGGCT 228
Qy 1864 CAGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1923
Db 229 CAGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 288
Qy 1924 TCACCAAGGCGAGCTGGAGACACCGGATCTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1983
Db 289 TCACCAAGGCGAGCTGGAGACACCGGATCTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 348
Qy 1984 GGATGCTGGGGGCTTCTGGGCTCAAGGAGGAGAACCCAGGAGCAAGGGGTTCCTCCAGGAGAA 2043
```

Db 2603 ATAGAGCCAGCTACCTACCTCTTCTCAGAGTGAAGAAATTAGAAATTTGTGATCATTC 2662
QY 2452 CTGTCCCAACATGGCTCCCGGGTATTCCTGGGCCACCTCGTCCGATAGGCCACAGAGGT 2511
Db 2663 CTGTCCCAACATGGCTCCCGGGTATTCCTGGGCCACCTCGTCCGATAGGCCACAGAGGT 2722
QY 2512 CCAGAGGATACCTGGTTTGCAGGAAGAGATGGTGTCTCTGGATTAGTGGTGTCCCT 2571
Db 2723 CCAGAGGATACCTGGTTTGCAGGAAGAGATGGTGTCTCTGGATTAGTGGTGTCCCT 2782
QY 2572 GGACGTCCAGAGTGTTCAGAGGATTAAGAGGCTTACCAGGAAGAAATGGGGAAGAGGAGC 2631
Db 2783 GGACGTCCAGAGTGTTCAGAGGATTAAGAGGCTTACCAGGAAGAAATGGGGAAGAGGAGC 2842
QY 2632 CAAGGTTTGGTATCTCTGGAGAACAGGTCTCTCTGGTCCCGAGGTCCAGAGGCGCT 2691
Db 2843 CAAGGTTTGGTATCTCTGGAGAACAGGTCTCTCTGGTCCCGAGGTCCAGAGGCGCT 2902
QY 2692 CTTGGAATAAGCAAGAGAGTCTCTCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGAC 2751
Db 2903 CTTGGAATAAGCAAGAGAGTCTCTCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGAC 2962
QY 2752 CATGGAATCCTGGATTCAGAGGCAACAGGCCCCCGAGGCATCTCGGACCCATCACTA 2811
Db 2963 CATGGAATCCTGGATTCAGAGGCAACAGGCCCCCGAGGCATCTCGGACCCATCACTA 3022
QY 2812 TGTCTTGTGTAATTCAGAGAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG 2865
Db 3023 TGTCTTGTGTAATTCAGAGAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG 3076

RESULT 5
AX430410
LOCUS AX430410 1184 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 192 from Patent WO0240715.
ACCESSION AX430410
VERSION AX430410.1 GI:21655774

KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1
AUTHORS Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.
TITLE Molecules for disease detection and treatment
JOURNAL Patent: WO 0240715-A 192 23-MAY-2002;
INCYTE GENOMICS INC (US)

FEATURES
Location/Qualifiers
1..1184
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Inclyce ID No: LI:108819.1:2000SEP08"

BASE COUNT 368 a 227 c 272 g 317 t
ORIGIN

Query Match 34.0%; Score 974; DB 6; Length 1184;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTCAGTATATACATTTCTCTGATGGTTTGGTGGTCTCTTTCAGAAATCTGTG 60
Db 211 ATGGCTCAGTATATACATTTCTCTGATGGTTTGGTGGTCTCTTTCAGAAATCTGTG 270
QY 61 TTAGCTGAAGATGGGAAGTATAGATCAAGTTCTCTGCTCTCTGCTCCGACAGATTTAGTTTC 120
Db 271 TTAGCTGAAGATGGGAAGTATAGATCAAGTTCTCTGCTCTCTGCTCCGACAGATTTAGTTTC 330
QY 121 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTGAAGATGTAAGAGTGGCTT 180
Db 331 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTGAAGATGTAAGAGTGGCTT 390
QY 181 GTCATATACAAAAAATCTTGACATAGGCGGCAAGTTTATTCAAGTTGGAGTGGTTCAA 240

Db 391 GTCAATATCAAAAAAATTGTGATAGGCCCAAGTTTATTCAAGTTGGATGGTTCAA 450
QY 241 TATAGTGAATACCTCTGTCTGGAGATTCCTCTCGGAAGCTATGATTTCAGGAGAAATTTG 300
Db 451 TATAGTGAATACCTCTGTCTGGAGATTCCTCTCGGAAGCTATGATTTCAGGAGAAATTTG 510
QY 301 AGGAGAGTGGGAATCCATCTACTTAGGAGGAAACACAAAGACAGGGAAGGCGATC 360
Db 511 AGGAGAGTGGGAATCCATCTACTTAGGAGGAAACACAAAGACAGGGAAGGCGATC 570
QY 361 CAGTTTGGCTCGATTAACCTTTTGGCAAGTCTCTCAGATTTCTGACTAAGATAGCAGTG 420
Db 571 CAGTTTGGCTCGATTAACCTTTTGGCAAGTCTCTCAGATTTCTGACTAAGATAGCAGTG 630
QY 421 GTACTTAGGATGGCAATCCCAAGATGAGCTCAAGGATGCAAGCTCAAGCAGCAAGAT 480
Db 631 GTACTTAGGATGGCAATCCCAAGATGAGCTCAAGGATGCAAGCTCAAGCAGCAAGAT 690
QY 481 AGTGAATACATTAATTTGGTATTTGGTGGTTTGCAGAAACAGAACTCCGAATCTAGA 540
Db 691 AGTGAATACATTAATTTGGTATTTGGTGGTTTGCAGAAACAGAACTCCGAATCTAGA 750
QY 541 GCTATTCGCAACAGGCTTCTCTACTTATGTTTATGTTGGAAGACTATATTCGCAATA 600
Db 751 GCTATTCGCAACAGGCTTCTCTACTTATGTTTATGTTGGAAGACTATATTCGCAATA 810
QY 601 TCCAAATAGGGAAGTATGACGACGAAACTTTGCGAAGATCTCTGTCGCAACAGCA 660
Db 811 TCCAAATAGGGAAGTATGACGACGAAACTTTGCGAAGATCTCTGTCGCAACAGCA 870
QY 661 ATTCAGTGGCAGCTCTGATGAAGGGGATTTGATATTTCTTTAGGTTTAGATGTAAT 720
Db 871 ATTCAGTGGCAGCTCTGATGAAGGGGATTTGATATTTCTTTAGGTTTAGATGTAAT 930
QY 721 AAAAGGTTAAGAAAGATACAGCTTTCACCAAAAAGATTAAGAGATAGAGTAACA 780
Db 931 AAAAGGTTAAGAAAGATACAGCTTTCACCAAAAAGATTAAGAGATAGAGTAACA 990
QY 781 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTTCACCAAGGCTCTCTTCATCA 840
Db 991 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTTCACCAAGGCTCTCTTCATCA 1050
QY 841 TATGATTTGTCTCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGAAATA 900
Db 1051 TATGATTTGTCTCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGAAATA 1110
QY 901 TTAATCTTGTGGAAGGCCAATAATAGCAGTTTACCTTTAAATGGTGTGGACAAAATCTTA 960
Db 1111 TTAATCTTGTGGAAGGCCAATAATAGCAGTTTACCTTTAAATGGTGTGGACAAAATCTTA 1170
QY 961 TTAATCTTGTGGAAGGCCAATAATAGCAGTTTACCTTTAAATGGTGTGGACAAAATCTTA 1170
Db 1171 TTAATCTTGTGGAAGGCCAATAATAGCAGTTTACCTTTAAATGGTGTGGACAAAATCTTA 1170

RESULT 6
AX430281
LOCUS AX430281 1169 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 63 from Patent WO0240715.
ACCESSION AX430281
VERSION AX430281.1 GI:21655645
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.
TITLE Molecules for disease detection and treatment
JOURNAL Patent: WO 0240715-A 63 23-MAY-2002;
INCYTE GENOMICS INC (US)
FEATURES Location/Qualifiers


```
QY 2092 GGAAATCAAGGTGAAAGAGGTATTACAGGTCCTCAAGAGGAGAGAAATGGAAGACAGGGAATT 2151
DB 2303 GGAAATCAAGGTGAAAGAGGTATTACAGGTCCTCAAGAGGAGAGAAATGGAAGACAGGGAATT 2362
QY 2152 CCAGGGCAACAGGGAATTCAGAGCCATCATCGTGCAGAAAGAGAGAGAGGTGAAAGGGA 2211
DB 2363 CCAGGGCAACAGGGAATTCAGAGCCATCATCGTGCAGAAAGAGAGAGAGGTGAAAGGGA 2422
QY 2212 GAACCTGTGTCCGAGGTGCAATTCGATCAAAAGAGAGAACTCGGGTGGATGGCTTGATG 2271
DB 2423 GAACCTGTGTCCGAGGTGCAATTCGATCAAAAGAGAGAACTCGGGTGGATGGCTTGATG 2482
QY 2272 GGGCCCGCAGGTCTCAAGGGCAACCTCGGGATCCAGTCTCCAGGACCCCCAGGTTTG 2331
DB 2483 GGGCCCGCAGGTCTCAAGGGCAACCTCGGGATCCAGTCTCCAGGACCCCCAGGTTTG 2542
QY 2332 GATGGGAAGCCCGAAGAGAGTTTCAGAACTTATTGACAAAGTTTCCACAGATGA 2391
DB 2543 GATGGGAAGCCCGAAGAGAGTTTTCAGAACTTATTGACAAAGTTTCCACAGATGA 2602
QY 2392 ATAAGAGCCAGCTACCACTGCTTACTTCAGAGTGGAGAAATAGAAATTTGTGATCATTCG 2451
DB 2603 ATAAGAGCCAGCTACCACTGCTTACTTCAGAGTGGAGAAATAGAAATTTGTGATCATTCG 2662
QY 2452 CTGTCCCAACATGCTCCCGGTTATCTCGGCACTCTGTCTCGATAGGCCAGAGGGT 2511
DB 2663 CTGTCCCAACATGCTCCCGGTTATCTCGGCACTCTGTCTCGATAGGCCAGAGGGT 2722
QY 2512 CCCAGAGATTACCTGTTTCCAGGAAGAGATGTTCTCTGATTTAGTGGGTGTCCT 2571
DB 2723 CCCAGAGATTACCTGTTTCCAGGAAGAGATGTTCTCTGATTTAGTGGGTGTCCT 2782
QY 2572 GGACCTCCAGGTGTCAGAGGATTAAGAGCCCTACCAAGAGAAATGGGGAAAAGGAGC 2631
DB 2783 GGACCTCCAGGTGTCAGAGGATTAAGAGCCCTACCAAGAGAAATGGGGAAAAGGAGC 2842
QY 2632 CAAGGTTTGGTATCTCGGAGACAGAGTCTCTGTCCTCCAGGTCAGAGGGCCCT 2691
DB 2843 CAAGGTTTGGTATCTCGGAGACAGAGTCTCTGTCCTCCAGGTCAGAGGGCCCT 2902
QY 2692 CTTGGAATAAGCAAGAGAGTCTCTCCAGGAGACCCAGGTCTCTCCCTGGCAAGATGGAGAC 2751
DB 2903 CTTGGAATAAGCAAGAGAGTCTCTCCAGGAGACCCAGGTCTCTCCCTGGCAAGATGGAGAC 2962
QY 2752 CATGAAACCTGGAATCAAGGGCAACAGGCCCCCAGGCATCTCGACCCATCACTA 2811
DB 2963 CATGAAACCTGGAATCAAGGGCAACAGGCCCCCAGGCATCTCGACCCATCACTA 3022
QY 2812 TGTTTGTGTAATTCGCAAGAGATCCGTTTCAGAAAGGACCAAACTATTAG 2865
DB 3023 TGTTTGTGTAATTCGCAAGAGATCCGTTTCAGAAAGGACCAAACTATTAG 3076
```

```
RESULT 4
HSM801597
LOCUS HSM801597 4160 bp mRNA linear PRI 20-MAR-2002
DEFINITION Homo sapiens mRNA; cDNA DKFZp564B052 (from clone DKFZp564B052);
complete cds.
ACCESSION AL336624
VERSION AL336624.1 GI:12052773
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4160)
AUTHORS Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and
Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wie-ann@dkfz-heidelberg.de;
```

sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp564B052) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clonedrzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cdna/>.

```
FEATURES
    Location/Qualifiers
        source          1..4160
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /map="6"
            /clone="DKFZp564B052"
            /tissue_type="brain"
            /clone_lib="564 (synonym: hfbz2). Vector pAMP1; host
            Xl-2blue; sites NotI + SalI"
            /dev_stage="fetal"
        gene            1..4160
            /gene="DKFZp564B052"
        CDS             203..3076
            /gene="DKFZp564B052"
            /note="dJ708P5.1, differences to genmodel, similarity to
            collagen proteins"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="CAB66559.1"
            /db_xref="GI:12052774"
            /db_xref="SPTREMBL:Q9H0V3"
            /translation="MAHYITFLCMVLVLLQNSVLAEDGEVRSRSCRTAPTDLVILDG
            SYSGPENFEIVRKWLVNIDFGPKFIQGVWQVSDYFVLEIPUGSYDSGEHLTA
            AVESILYLGNTKTKGAIQALDYLDPKSRFLTKIAVLTDGKSQDDVKDAQAARD
            SKILFLAIGVSTEDAEALAIANKPSSTVFYVEDYIAISKIREVYKQKLCEESVCP
            TRIPVAARDERGFIDLLGVNKKRIOLSPKIKGYEVTSKVLSELTSGVFBEG
            LPPSYVFVSTORFKVKIIDLWLRLITDGRPOLAVTLNGVDKILLFTTTSVINGSOV
            TFANPVKTLFDSEWHQIRLLVTEQDVTLYIDQOQIENKPLHPVLGILLINGTOIGKY
            SGKSETVQFVQKLRIYCDPEQNNRETACIPGNGECLNGSPDSVSTPAPCLCPGK
            PGLQPGKDPGLPQNGPQGDGKPGYGGIAGTPGVPGSPGIQAGRGLPQYKGE
            GRAGEKDRGLPGLGHPMGSGKGAAGDKSPGFYKGAKEKGNAGPGLPG
            PGPFGSHGKDGGLMGSPFGKEAGSPAGQDGTGEGFIPGPNRGLMGKGLG
            PQOQKKGAPMGLMQSGSPQPGT*PGSKSGKSPGPG*QMPGAGLKGEPGATGSP
            GEPYMGLPGLTGKKGKNGKEKGLCGOKGNGRGI*PGOQGIQSHHAKGERGSKG
            E2GVRGALSGKSGSGVGLMGAPAGKQPGQDPGFPQSPDLGXPRFESQFIRQVCT
            DVIRAQLPVLQJQSGIRNRNCHLSQHGSPGIPGPPGP*GEPGPRG*PGLPGRDGVPL
            VGPVPRGVRLKGLPGRNGEKSGQGFYGEQGGPFGPFGPGPGGPGISKEGPPGDP
            PGKGDGHKFG*QQQPGPGICDPSLCFSV*ARRDPFRKGPY"
        polyA_signal     4120..4125
            /gene="DKFZp564B052"
        polyA_site       4141
            /gene="DKFZp564B052"
BASE COUNT 1271 a 795 c 1026 g 1068 t
ORIGIN
```

```
Query Match      71.9%; Score 2061; DB 9; Length 4160;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2861; Conservative 0; Mismatches 4; Indels 9; Gaps 1;

QY 1 ATGGCTCACTATATACATTCTCTGCATCGTTTGGTCTGCTTCTTCAGAAATTCGTG 60
DB 203 ATGGCTCACTATATACATTCTCTGCATCGTTTGGTCTGCTTCTTCAGAAATTCGTG 262
QY 61 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGTCGTACTGCTCCGACAGATTAGTTTTC 120
DB 263 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGTCGTACTGCTCCGACAGATTAGTTTTC 322
QY 121 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTTGAAATAGT*GAAAAAGTGGCTT 190
DB 323 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTTGAAATAGT*GAAAAAGTGGCTT 382
QY 131 GTCAATATCACAAAAAATCTTGACATAGGCCGAAAGTTTATTCAGTTGGAGTGGTTCAA 240
DB 383 GTCAATATCACAAAAAATCTTGACATAGGCCGAAAGTTTATTCAGTTGGAGTGGTTCAA 442
```

ORIGIN

Query Match									
Best Local Similarity 71.9%; Score 2061; DB 6; Length 4160;									
Matches 2861; Conservative 0; Mismatches 4; Indels 9; Gaps 2;									
Qy	1	ATGGCTCACTATATTACATTTCTCGCATGGTTTGGTGGCTCTCTTCAAAATTCGTG	60						
Db	203	ATGGCTCACTATATTACATTTCTCGCATGGTTTGGTGGCTCTCTTCAAAATTCGTG	262						
Qy	61	TTAGCTGAAGTGGGAACTAGATCAAGTTCGTACTGCTCCGACAGATTTAGTTTC	120						
Db	263	TTAGCTGAAGTGGGAACTAGATCAAGTTCGTACTGCTCCGACAGATTTAGTTTC	322						
Qy	121	ATCTTAGTGGCTCTTATAGTGTGGCCGAGAAACCTTGAATAGTGAAGTGGCTT	180						
Db	323	ATCTTAGTGGCTCTTATAGTGTGGCCGAGAAACCTTGAATAGTGAAGTGGCTT	382						
Qy	181	GTCAATATCACAAAACCTTTCACATAGGGCCGAGATTTATTCAGTTTGGAGTGGTCAA	240						
Db	383	GTCAATATCACAAAACCTTTCACATAGGGCCGAGATTTATTCAGTTTGGAGTGGTCAA	442						
Qy	241	TATAGTGAATACCTGTGTGGAGATTCCTCTCGAAGCTATGATTCAGGAGAACATTTG	300						
Db	443	TATAGTGAATACCTGTGTGGAGATTCCTCTCGAAGCTATGATTCAGGAGAACATTTG	502						
Qy	301	ACGGCAGCAGTGGAAATCCATCTCTACTTTAGGAGGAAACACAAAGACAGGAGGCCATC	360						
Db	503	ACGGCAGCAGTGGAAATCCATCTCTACTTTAGGAGGAAACACAAAGACAGGAGGCCATC	562						
Qy	361	CAGTTTGCCTCGATTTACCTTTTGGCAAGTCTCTCAGATTTCTGACTAAGATAGCAGTG	420						
Db	563	CAGTTTGCCTCGATTTACCTTTTGGCAAGTCTCTCAGATTTCTGACTAAGATAGCAGTG	622						
Qy	421	GTACTTACGATGGCAATCCCAAGATGAGCTCAAGATGAGCTCAAGCAGCAGAGAT	480						
Db	623	GTACTTACGATGGCAATCCCAAGATGAGCTCAAGATGAGCTCAAGCAGCAGAGAT	682						
Qy	481	AGTAAGATAACATTTATTTGCTATTTGGTGTGGTTCAGAAACAGAAAGATGCCGAACTTGA	540						
Db	683	AGTAAGATAACATTTATTTGCTATTTGGTGTGGTTCAGAAACAGAAAGATGCCGAACTTGA	742						
Qy	541	GCTATTGCCAAACAGCCCTCGTCTACTTATGTGTTTATGTGGAGAGACTATATTGCAATA	600						
Db	743	GCTATTGCCAAACAGCCCTCGTCTACTTATGTGTTTATGTGGAGAGACTATATTGCAATA	802						
Qy	601	TCCAAAATAAGGGAAGTGAAGAGCAAACTTTGTGAAGATCTGTCTGTCCTCAACACGA	660						
Db	803	TCCAAAATAAGGGAAGTGAAGAGCAAACTTTGTGAAGATCTGTCTGTCCTCAACACGA	862						
Qy	661	ATTCAGTGGCAGCTCGTGAATGAAGGGGATTTGATTTCTTTTAGTTTATAGTGTAAAT	720						
Db	863	ATTCAGTGGCAGCTCGTGAATGAAGGGGATTTGATTTCTTTTAGTTTATAGTGTAAAT	922						
Qy	721	AAAAAGTTAAGAAAGAAATACAGCTTTCACCAAAAAGATAAAGGATATGAGTAACA	780						
Db	923	AAAAAGTTAAGAAAGAAATACAGCTTTCACCAAAAAGATAAAGGATATGAGTAACA	982						
Qy	781	TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTCCCGAAGGCTTCTCTCCATCA	840						
Db	983	TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTCCCGAAGGCTTCTCTCCATCA	1042						
Qy	841	TATGTATTTGTGCTACTCAAGATTTTAAAGTCAAGAAATTTGGATTTATGGAGATA	900						
Db	1043	TATGTATTTGTGCTACTCAAGATTTTAAAGTCAAGAAATTTGGATTTATGGAGATA	1102						
Qy	901	TTAACTATTGATCGAAGGCCACAAATAGCAGTTTACCTTAAATGGTGTGGACAAATCTTA	960						
Db	1103	TTAACTATTGATCGAAGGCCACAAATAGCAGTTTACCTTAAATGGTGTGGACAAATCTTA	1162						
Qy	961	TTATTTACAAACACAGCGTAATTAATGGCTCAAGTGGTTACCTTTGCTAACCTCAA	1020						
Db	1163	TTATTTACAAACACAGCGTAATTAATGGCTCAAGTGGTTACCTTTGCTAACCTCAA	1222						

Qy	1201	GTAAAGACCTTTGTTGATGAAGGCTGGCACCACAAATTCGTCTCTTAGTAACAGAACAGAT	1280
Db	1223	GTAAAGACCTTTGTTGATGAAGGCTGGCACCACAAATTCGTCTCTTAGTAACAGAACAGAT	1282
Qy	1081	GTGACTTTGTATATATGATGACCAACAAATTCGAAACAGCCCTTACATCCAGTTTAAAGG	1140
Db	1283	GTGACTTTGTATATATGATGACCAACAAATTCGAAACAGCCCTTACATCCAGTTTAAAGG	1342
Qy	1141	ATCTTATGATCAATGGGCAACCCAAATTCGAAACAAATTCGAAACAGCCCTTACATCCAG	1200
Db	1343	ATCTTATGATCAATGGGCAACCCAAATTCGAAACAAATTCGAAACAGCCCTTACATCCAG	1402
Qy	1201	TTTGTATGTCACAAAGCTTGGCAATCTACTGTGACCCAGAACAGAACACCCGGGACAGCA	1260
Db	1403	TTTGTATGTCACAAAGCTTGGCAATCTACTGTGACCCAGAACAGAACACCCGGGACAGCA	1462
Qy	1261	TGTGAGATTCCTGGATTT-----TSCCTTAATGTCTCCAGTGAATAGGTCAACT	1311
Db	1463	TGTGAGATTCCTGGATTTAATGGAGAGTGGCTTAATGGTCCAGTGAATAGGTCAACT	1522
Qy	1312	CCAGCTCCCTGTATTTTGTCTCCCGGAAACACAGACATTCAGGCCCCAAAGGTGACCT	1371
Db	1523	CCAGCTCCCTGTATTTTGTCTCCCGGAAACACAGACATTCAGGCCCCAAAGGTGACCT	1582
Qy	1372	GGACTGCCCTGGAAACCCCTGGCTACCCCTGGACAACTGGTCAAGATGGTAAGCCTGGATAT	1431
Db	1583	GGACTGCCCTGGAAACCCCTGGCTACCCCTGGACAACTGGTCAAGATGGTAAGCCTGGATAT	1642
Qy	1432	CAGGGAATTTGAGGAGACACAGGTGTTCCAGGATCTCCAGGAATACAAAGAGCTCGAGGA	1491
Db	1643	CAGGGAATTTGAGGAGACACAGGTGTTCCAGGATCTCCAGGAATACAAAGAGCTCGAGGA	1702
Qy	1492	CTACCAAGTTTCAAAAGGAGAACCCAGGCGAGATGTCGACAGGCTGATCGTGGACTTCCT	1551
Db	1703	CTACCAAGTTTCAAAAGGAGAACCCAGGCGAGATGTCGACAGGCTGATCGTGGACTTCCT	1762
Qy	1552	GGTTTTCCTGGCTTCATGGCATGCCAGGATCAAAAGGTGAAATGGGTGCCAAAGGAGAC	1611
Db	1763	GGTTTTCCTGGCTTCATGGCATGCCAGGATCAAAAGGTGAAATGGGTGCCAAAGGAGAC	1822
Qy	1612	AAAGGATCACCTGGATTTTATGGCAAAAGGTGCAAAAGGTGCAAAAGGGAATGCTGGC	1671
Db	1823	AAAGGATCACCTGGATTTTATGGCAAAAGGTGCAAAAGGTGCAAAAGGGAATGCTGGC	1882
Qy	1672	TTCCCTGGCTCCCTGGACCTGCTGGAGAACCCAGGAGACATGGAAAGGATGGAATATG	1731
Db	1883	TTCCCTGGCTCCCTGGACCTGCTGGAGAACCCAGGAGACATGGAAAGGATGGAATATG	1942
Qy	1732	GGTAGTCCCGTTTCAAGGGAGAACCCAGGATTCCTGCTGCTCCCGGGCAGATGGAACA	1791
Db	1943	GGTAGTCCCGTTTCAAGGGAGAACCCAGGATTCCTGCTGCTCCCGGGCAGATGGAACA	2002
Qy	1792	CGGGAGAGCCCTGGAAATCCAGGATTTCTGGAACCCAGGATTAATGGGCAAAAGGGA	1851
Db	2003	CGGGAGAGCCCTGGAAATCCAGGATTTCTGGAACCCAGGATTAATGGGCAAAAGGGA	2062
Qy	1852	GAATTTGGCTCCAGGACAGCAAGGAAAGGAGCCCGAGGATGCTGTTTAAATG	1911
Db	2063	GAATTTGGCTCCAGGACAGCAAGGAAAGGAGCCCGAGGATGCTGTTTAAATG	2122
Qy	1912	GGAGCAATGGCTCACAGGCCAGCTCGGAACACCGGATCTAAGGGAAGCAAGGTAATG	1971
Db	2123	GGAGCAATGGCTCACAGGCCAGCTCGGAACACCGGATCTAAGGGAAGCAAGGTAATG	2182
Qy	1972	CCTGGAATTTCAAGGATGCTGGGCTTCTGGGCTCAAGGGAGAACCCAGGACACCGGT	2031
Db	2183	CCTGGAATTTCAAGGATGCTGGGCTTCTGGGCTCAAGGGAGAACCCAGGACACCGGT	2242
Qy	2032	TCCCAAGGAGAACCCAGGATACATGGTTTACCCGGATTCAGGAAAGGAGGAGACAA	2091
Db	2243	TCCCAAGGAGAACCCAGGATACATGGTTTACCCGGATTCAGGAAAGGAGGAGACAA	2302

1223	DB	GTTAAGACGTTGTTGATGAAGCGTCGCACCAATTTCTGCTCTTAGTAAAGAACAGAT	1282
1081	QY	GTGACTTTGTATATTGATGACCACAAATTTGAANAACAGGCCCTTACATCCAGATTTTAGGG	1140
1283	DB	GTGACTTTGTATATTGATGACCACAAATTTGAANAACAGGCCCTTACATCCAGTTTTAGGG	1342
1141	QY	ATCTTGATCAATTTGGGCAAAACCCAAATTTGGAATAATTCTTGAAAGAAGAAACTGTTGAG	1200
1343	DB	ATCTTGATCAATTTGGGCAAAACCCAAATTTGGAATAATTCTTGAAAGAAGAAACTGTTGAG	1402
1201	QY	TTTGTATGTCCAAAGTTGCGAATCTACTGTGACCCAGAACAGAACCAACCCGGGAGACAGCA	1260
1403	DB	TTTGTATGTCCAAAGTTTGCGAATCTACTGTGACCCAGAACAGAACCAACCCGGGAGACAGCA	1462
1261	QY	TGTGAGATTCTGTGGATT-----TGCCTTAATGTCCTCAGTGATGTAGGTTTCAACT	1311
1463	DB	TGTGAGATTCTGTGGATTAAATGGAGAGTGCCCTTAATGTTCCCAAGTGATGTAGGTTTCAACT	1522
1312	QY	CCAGCTCCCTGTATTTCTGCTCCGGGAACACAGAGACTTCAAGGCCCCCAAGAGTGACCCCT	1372
1523	DB	CCAGCTCCCTGTATTTGCTCTCGGGAACACAGAGACTTCAAGGCCCCCAAGAGTGACCCCT	1582
1372	QY	GGACTGCTGGGAACCCCTGCTACCTCGCAACCTGTGCAAGATGSGTAAGCCTGSGATAT	1431
1583	DB	GGACTGCTGGGAACCCCTGCTACCTCGCAACCTGTGCAAGATGSGTAAGCCTGSGATAT	1642
1432	QY	CAGGGAATTGACGGGACACAGGTGTTCCAGGATCTCCAGGNAATACAGGAGCTCGAGGA	1491
1643	DB	CAGGGAATTGACGGGACACAGGTGTTCCAGGATCTCCAGGNAATACAGGAGCTCGAGGA	1702
1492	QY	CTACCAAGTTACAAAGAGAAACAGAGCGCAGATGTTGACAAGGGTGATCTGTGGACTTCCT	1551
1703	DB	CTACCAAGTTACAAAGAGAAACAGAGCGCAGATGTTGACAAGGGTGATCTGTGGACTTCCT	1762
1552	QY	GGTTTTCTGGGCTTCATGGCATGCCAGGATCAAAGGGTGAAATGGGTGCCAAAGAGAGAC	1611
1763	DB	GGTTTTCTGGGCTTCATGGCATGCCAGGATCAAAGGGTGAAATGGGTGCCAAAGAGAGAC	1822
1612	QY	AAAGGATCACCTGGATTATTATGGCAAAAGGGTGCAAAAGGTGAAAAGGGCAATGCTGGC	1671
1823	DB	AAAGGATCACCTGGATTATTATGGCAAAAGGGTGCAAAAGGTGAAAAGGGCAATGCTGGC	1882
1672	QY	TTCCCTGGCCTCCCTGGACCTGCTGGAGAACCCAGGAAGACATGGAAGGATGGAATTAATG	1731
1883	DB	TTCCCTGGCCTCCCTGGACCTGCTGGAGAACCCAGGAAGACATGGAAGGATGGAATTAATG	1942
1732	QY	GGTAGTCCCGGTTTCAAGGGAGACAGGATCCCTGGTGCTCCGGGACGAGTGGGAACA	1791
1943	DB	GGTAGTCCCGGTTTCAAGGGAGACAGGATCCCTGGTGCTCCGGGACGAGTGGGAACA	2002
1792	QY	CGGGAGAGCCTGGAAATCCCAAGGATTTCTCGAAACCCAGGATTAATGGGCCAAAAGGGA	1851
2003	DB	CGGGAGAGCCTGGAAATCCCAAGGATTTCTCGAAACCCAGGATTAATGGGCCAAAAGGGA	2062
1852	QY	GAAATTGGGCTCCAGGACAGCAGGAAAAAAGGAGCCGAGGATGCTCTGTTTAATG	1911
2063	DB	GAAATTGGGCTCCAGGACAGCAGGAAAAAAGGAGCCGAGGATGCTCTGTTTAATG	2122
1912	QY	GGAAGCAATGGCTCACCAAGGCCAGCCTCGGAACACCGGGATCTTAAGGGGAAGCAAAAGTCAA	1971
2123	DB	GGAAGCAATGGCTCACCAAGGCCAGCCTCGGAACACCGGGATCTTAAGGGGAAGCAAAAGTCAA	2182
1972	QY	CCTGGNAATTCAAGGATGCTGGGCTCTTGAGGCTCAAGGGAGAACCCAGGAGCAACCGGT	2031
2183	DB	CCTGGNAATTCAAGGATGCTGGGCTCTTGAGGCTCAAGGGAGAACCCAGGAGCAACCGGT	2242
2032	QY	TCCCCAGGAGAACACAGGATACATGGGTTTACCCGGGATTCAGGAAAAAAGGGGACAAA	2091
2243	DB	TCCCCAGGAGAACACAGGATACATGGGTTTACCCGGGATTCAGGAAAAAAGGGGACAAA	2302
2092	QY	GGAATTAAGGTTGAAAAAGGATTACAGGCTCAAAAGGGAGAAAAATGGAGACAGGGAATT	2151

D5	2303	GGAAATCAAGGTGAAAAGGTGTTCTCAGGGTCAAAAGGGGAAAAATTTGGAAGACACGGAAATT	2362
Qy	2152	CCAGGGCAACAGGGAATTTCAAGGCCATCATGGTGCAAAAGGAGAGAGAGGTGAAAGGGA	2211
D5	2363	CCAGGGCAACAGGGAATTTCAAGGCCATCATGGTGCAAAAGGAGAGAGAGGTGAAAGGGA	2422
Qy	2212	GAACCTGGTGTCCGAGTGCATTTGGATCANAGAGAAATCTGGGGTCGATGGCTCATG	2271
D5	2423	GAACCTGGTGTCCGAGTGCATTTGGATCANAGAGAAATCTGGGGTCGATGGCTCATG	2482
Qy	2272	GGGCGCCGAGGTCCTTAAGGGGGCAACCTCGGGGATCCAGGTCCCTCAGGACCCGCCAGGTTTG	2331
D5	2483	GGGCGCCGAGGTCCTTAAGGGGGCAACCTCGGGGATCCAGGTCCCTCAGGACCCGCCAGGTTTG	2542
Qy	2332	GATGGGAAGCCGGAGAGAGTATTTTCAACAATTTATTTCGACAAGATTTGCACAGATGTA	2391
D5	2543	GATGGGAAGCCGGAGAGAGTATTTTCAACAATTTATTTCGACAAGATTTGCACAGATGTA	2602
Qy	2392	ATAAGAGCCCAAGCTACCAAGTCTTACTTTCAGAGTGGGAAGAAATTTAGAAATTTGTCATCATGTC	2451
D5	2603	ATAAGAGCCCAAGCTACCAAGTCTTACTTTCAGAGTGGGAAGAAATTTAGAAATTTGTCATCATGTC	2662
Qy	2452	CTGTCCCAACATGGCTCCCGGGTATTTCTTGGGCCACCTGGTCCGATAGGCCCGAGAGGGT	2511
D5	2663	CTGTCCCAACATGGCTCCCGGGTATTTCTTGGGCCACCTGGTCCGATAGGCCCGAGAGGGT	2722
Qy	2512	CCCAAGAGATTAACCTGGTTTGCCAGGAAGAGATGGTGTTCCTGGATTAAGTGGGTGTCCCT	2571
D5	2723	CCCAAGAGATTAACCTGGTTTGCCAGGAAGAGATGGTGTTCCTGGATTAAGTGGGTGTCCCT	2782
Qy	2572	GGACGTCCAGGTGTCCAGAGGATTTAAAAGGCCTACCAAGGAAGAAATTTGGGGAAAAAGGGAGC	2631
D5	2783	GGACGTCCAGGTGTCCAGAGGATTTAAAAGGCCTACCAAGGAAGAAATTTGGGGAAAAAGGGAGC	2842
Qy	2632	CAAGGGTTTGGGTATCTGTGAGAACAGGTCTCTGTGTCCTCCCGCCAGGTCCAGAGGGCCCT	2691
D5	2843	CAAGGGTTTGGGTATCTGTGAGAACAGGTCTCTGTGTCCTCCCGCCAGGTCCAGAGGGCCCT	2902
Qy	2692	CCTGGAAATGAAGCAAGAGGTCTCTCCAGGAGAGCCCAAGTCTCCCTTGGCAAAAGATGGAGAC	2751
D5	2903	CCTGGAAATGAAGCAAGAGGTCTCTCCAGGAGAGCCCAAGTCTCCCTTGGCAAAAGATGGAGAC	2962
Qy	2752	CATGGAACCTGGATTCCTCAAGGGCAACCAAGGCCCCCAAGGCATCTCGACCCATCACTA	2811
D5	2963	CATGGAACCTGGATTCCTCAAGGGCAACCAAGGCCCCCAAGGCATCTCGACCCATCACTA	3022
Qy	2812	TGTTTTAGTGTAAATTTGCCAGAAGAGATCCGTTTCAGAAAAAGGACCAAACTATTAG	2865
D5	3023	TGTTTTAGTGTAAATTTGCCAGAAGAGATCCGTTTCAGAAAAAGGACCAAACTATTAG	3076

```

RESULT 3
AX086127 LOCUS 4160 bp DNA linear PAT 09-MAR-2001
DEFINITION Sequence 79 from Patent WO0112659.
ACCESSION AX086127
VERSION AX086127.1 GI:13275878
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Wiemann,S.
AUTHORS Human dna sequences
TITLE Patent: WO 0112659-A 79 22-FEB-2001;
JOURNAL German Human Genome Project (DE)
FEATURES
Location/Qualifiers
source 1..4160
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 1271 a 795 c 1026 g 1068 t

```


Qy 2701 AGCAAGAGGTCTCTCAGGAGACCCAGGTCTCCCTGGCAAGATGGAGACCATGAA 2760
 Db 2863 AGCAAGAGGTCTCTCAGGAGACCCAGGTCTCCCTGGCAAGATGGAGACCATGAA 2922
 Qy 2761 CCTGGAATCCAGGAGACCCAGGCCCCCAGGAGCATCTGGAGCCATCATATGTTAGT 2820
 Db 2923 CCTGGAATCCAGGAGACCCAGGCCCCCAGGAGCATCTGGAGCCATCATATGTTAGT 2982
 Qy 2821 GTAATGCCAGAGATCCGTTTCAGAAAGAGACCAAACTATTAG 2865
 Db 2983 GTAATGCCAGAGATCCGTTTCAGAAAGAGACCAAACTATTAG 3027

RESULT 2
 AF414088 4160 bp mRNA linear PRI 16-NOV-2001
 LOCUS Homo sapiens collagen XXI (COL21A1) mRNA, complete cds.
 DEFINITION
 ACCESSION AF414088
 VERSION AF414088.1 GI:15593269

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 (bases 1 to 4160)
 TITLE Fitzgerald, J. and Bateman, J. F.
 JOURNAL A new FACIT of the collagen family. COL21A1
 MEDLINE PUBMED 21566190

REFERENCE
 2 (bases 1 to 4160)
 TITLE Fitzgerald, J. and Bateman, J. F.
 JOURNAL Direct Submission
 AUTHORS Submitted (23-AUG-2001) Cell and Matrix Biology, Murdoch Children's
 JOURNAL Research Institute, Flemington Rd, Parkville, Vic 3052, Australia

FEATURES
 source
 1. .4160
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="6p12-11"
 1. .4160
 /gene="COL21A1"
 203. .3076
 /gene="COL21A1"
 /note="FACIT collagen: similar to protein encoded by
 GenBank Accession Number ALI36624"
 /codon_start=1
 /product="collagen XXI"
 /protein_id="AA02227.1"
 /db_xref="GI:15593270"
 /translations="MAHYITFLCMLVLLONSVALEDEVRSSRTAPTDLVFILOG
 SYSVGPENFEIVKMLVNTTKNFDIGPKFIOGVVQYSDYVPLEIPGSDVSGEHLTA
 AVESILYLGNTKGTALIOFALDYLFAKSSRELTIAVLTQKSDQAVKDAQNAAD
 SKITLFAIYGVSTEDAELEATANKPSSYFVVEDYIAISKIREVMQKLCESVCP
 TRIPVAARDERGDFILGLDGVNKKVKKRIQLSPKIKIGFEVTSKVLSELTSNVFPEG
 LPSYVYFVSTPRFKKKIIDLRLITIDRPQIAVTLNMGVDKILLFTTTSVINGSQV
 TFEANQVTLFDGMHQIRLVLEQDVTLYIDQQIENKPLHVLGILINGICQIGKY
 SKKEETVDFQVKLRYCDPEQNNRETACEIPFNCECLNGSPDVGSTPAPICPGPK
 PGLQPKGDLPLGNPYPGQCDGKPGYQGIAGTPGVPGSPGIQAGALPGYKGP
 GRDGDGDRGLPGNPGHMGPSKSGEMAKGDKGSPGVKKGAKGKGNAGPGLPG
 PAGEPGRHKGDLGLMGSPFKGEGSPGAPQDGTGRGFIQFPNRLQMLQKGEIGP
 PGQGGKGAQPMGLMNGSPGPGTIPGSKSGKEPGIQGMPGASGLUKGEPGATGP
 GEGPYMGLPQIQKDKGNQGEKGIQGGNGRQGIIPGQGGIQGHGAKGERGEK
 EPGVRAIGLSKGSVDGLMGAPGPKGPDGPGQPLDGLDKGRFSEQFIHQVCT
 DVIRAGLVLLQSGRIRNCDHCLSLSHGSPGIPGPFPIGPEGRGLPGLGRDGVPLG
 VGVPRGPRVRLKGLPGRNGEKGSGQGFYVPGQGPGRPGPGPGIISKEGPPGDPGL
 PGKDGPHGKPGIQGQPGPGICDPSLCFSVIARRDPFRKPNY"

BASE COUNT
 1270 a 796 c 1026 g 1068 t
 ORIGIN

Query Match 73.7%; Score 2112; DB 9; Length 4160;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 2862; Conservative 0; Mismatches 3; Indels 5; Gaps 1;
 Qy 1 ATGSGCTCAGCTATATACATTTCTCTGCATGGTTTGGTGGCTGCTTCTTCAGAAATCTGTG 60
 Db 203 ATGGCTCAGCTATATACATTTCTCTGCATGGTTTGGTGGCTGCTTCTTCAGAAATCTGTG 262
 Qy 61 TTAGCTGAAGATGGGGAGTAAGATCAAGTTGTCTACTCTCCGACAGATTTTAAAGTTTC 120
 Db 263 TTAGCTGAAGATGGGGAGTAAGATCAAGTTGTCTACTCTCCGACAGATTTTAAAGTTTC 322
 Qy 121 ATCTTAGATGGCTCTTATAGTCTTGGCCAGAAAACCTTTGAAATAGTGAAGATGGCTT 180
 Db 323 ATCTTAGATGGCTCTTATAGTCTTGGCCAGAAAACCTTTGAAATAGTGAAGATGGCTT 382
 Qy 181 GTCATATACAAAAAATTTTGCATAGGGCCGAAGTTTATTTCAAGTTGGAGTGGTCAA 240
 Db 383 GTCATATACAAAAAATTTTGCATAGGGCCGAAGTTTATTTCAAGTTGGAGTGGTCAA 442
 Qy 241 TATAGTGAATACCTCTGTCTGGAGATTCCTCTCGAAGCTATGATTCAGAGAGACATTTG 300
 Db 443 TATAGTGAATACCTCTGTCTGGAGATTCCTCTCGAAGCTATGATTCAGAGAGACATTTG 502
 Qy 301 ACGGCAGCAGTGGAAATCCATCTACTTAGAGGAAACACAAAGACAGAGAGAGCCATC 360
 Db 503 ACGGCAGCAGTGGAAATCCATCTACTTAGAGGAAACACAAAGACAGAGAGAGCCATC 562
 Qy 361 CAGTTTGGCTCGATTACCTTTTCCAAAGTCCCTACGATTTCTGACTTAAGATAGCAGT 420
 Db 563 CAGTTTGGCTCGATTACCTTTTCCAAAGTCCCTACGATTTCTGACTTAAGATAGCAGT 622
 Qy 421 GTACTTACGATGGCAATCCCAAGATGACGCTCAAGGATGAGCTCAAGAGCAAGAGAT 480
 Db 623 GTACTTACGATGGCAATCCCAAGATGACGCTCAAGGATGAGCTCAAGAGCAAGAGAT 682
 Qy 481 AGTAAGATAACATTATTTGCTATTGCTGTGGTTCAGAAAACAGAGATGCCAACTTAGA 540
 Db 683 AGTAAGATAACATTATTTGCTATTGCTGTGGTTCAGAAAACAGAGATGCCAACTTAGA 742
 Qy 541 GCTATTGCCAACAAGGCTTCGCTACTTATGTTTATGTTTATGTTGGAAGACTATATTGCAATA 600
 Db 743 GCTATTGCCAACAAGGCTTCGCTACTTATGTTTATGTTTATGTTGGAAGACTATATTGCAATA 802
 Qy 601 TCCAAAATTAAGGGAAGTATGATGAAGCAGAAACCTTTGTGAAGATCTGTCTGCCAACAGA 660
 Db 803 TCCAAAATTAAGGGAAGTATGATGAAGCAGAAACCTTTGTGAAGATCTGTCTGCCAACAGA 862
 Qy 661 ATTCCAGTGGCAGCTCGTGATGAAGGGGATTTGATATCTTTTAGGTTTAGATGTAAT 720
 Db 863 ATTCCAGTGGCAGCTCGTGATGAAGGGGATTTGATATCTTTTAGGTTTAGATGTAAT 922
 Qy 721 AAAAAGCTTAAGAAAGATACAGCTTTCACCAAAAAGATAAAGGATATGAAGTAAACA 780
 Db 923 AAAAAGCTTAAGAAAGATACAGCTTTCACCAAAAAGATAAAGGATATGAAGTAAACA 982
 Qy 781 TCAAAGTTGATTTATCAGAACTCAAGCAATGTTTCCAGAAAGTCTTCTCCATCA 840
 Db 983 TCAAAGTTGATTTATCAGAACTCAAGCAATGTTTCCAGAAAGTCTTCTCCATCA 1042
 Qy 841 TATGTTATTGCTCTACTCAAGATTTTAAAGTCAAGAAAATTTGGGATTTATGGAGATA 900
 Db 1043 TATGTTATTGCTCTACTCAAGATTTTAAAGTCAAGAAAATTTGGGATTTATGGAGATA 1102
 Qy 901 TTAATCTATTGATGGAGGCCCAAAATAGCAGTTTAAATGGTGTGGACAAAATCTTA 960
 Db 1103 TTAATCTATTGATGGAGGCCCAAAATAGCAGTTTAAATGGTGTGGACAAAATCTTA 1162
 Qy 961 TTAATCTAAGAACACAGCGTATTAATGGCTCACAAGTGGTTTACCTTTGCTAACCTCAA 1020
 Db 1163 TTAATCTAAGAACACAGCGTATTAATGGCTCACAAGTGGTTTACCTTTGCTAACCTCAA 1222
 Qy 1021 GTTAAGACGTTGTTTGTATGAAGGCTGGCACCACAAATTCGTCTCTTAGTAAACAGAACAGAT 1080

C 66	21	0.7	150236	2	AC023292	AC023292 Homo sapi
C 67	21	0.7	157711	9	AC055738	AC055738 Homo sapi
C 68	21	0.7	159160	2	AC024715	AC024715 Homo sapi
C 69	21	0.7	169986	2	AL353589	AL353589 Homo sapi
C 70	21	0.7	172506	2	AC097010	AC097010 Sus scrofa
C 71	21	0.7	175099	2	AC017029	AC017029 Homo sapi
C 72	21	0.7	177476	2	AC114629	AC114629 Mus muscu
C 73	21	0.7	178033	9	AC107025	AC107025 Homo sapi
C 74	21	0.7	178415	2	AC023547	AC023547 Homo sapi
C 75	21	0.7	179523	2	AC026104	AC026104 Homo sapi
C 76	21	0.7	179854	2	AL143471	AL143471 Macaca mu
C 77	21	0.7	183104	9	AL355601	AL355601 Human DNA
C 78	21	0.7	186830	2	AC121539	AC121539 Mus muscu
C 79	21	0.7	188247	9	AL442838	AL442838 Homo sapi
C 80	21	0.7	190394	9	AL578934	AL578934 Human DNA
C 81	21	0.7	191463	9	AC078814	AC078814 Homo sapi
C 82	21	0.7	196127	2	AC132788	AC132788 Rattus no
C 83	21	0.7	203978	2	AC109286	AC109286 Mus muscu
C 84	21	0.7	207945	9	CNS01055	AL121841 Human chr
C 85	21	0.7	208026	9	AP006292	AP006292 Homo sapi
C 86	21	0.7	213082	2	AC130126	AC130126 Rattus no
C 87	21	0.7	219184	2	AC132830	AC132830 Mus muscu
C 88	21	0.7	219358	2	AC135740	AC135740 Rattus no
C 89	21	0.7	220688	2	AC116463	AC116463 Mus muscu
C 90	21	0.7	234549	2	AC105497	AC105497 Rattus no
C 91	21	0.7	251244	2	AC103149	AC103149 Rattus no
C 92	21	0.7	286957	2	AC131144	AC131144 Rattus no
C 93	20	0.7	393	6	AX507730	AX507730 Sequence
C 94	20	0.7	400	11	GI6945	GI6945 human STS S
C 95	20	0.7	448	9	AF463542	AF463542 Gorilla g
C 96	20	0.7	548	6	AR225522	AR225522 Sequence
C 97	20	0.7	548	6	AX321643	AX321643 Sequence
C 98	20	0.7	637	11	BV000957	BV000957 S208P6187
C 99	20	0.7	954	6	AR310334	AR310334 Sequence
C 100	20	0.7	954	6	AX097721	AX097721 Sequence

ALIGNMENTS

RESULT 1	AF330693	4122 bp	mRNA	linear	PR: 01-MAR-2002
LOCUS	AF330693	Homo sapiens alpha 1 chain-like collagen COL1A1 precursor (COL1A1)			
DEFINITION	AF330693	mRNA, complete cds.			
ACCESSION	AF330693				
VERSION	AF330693.1	GI:17974509			
KEYWORDS					
SOURCE		Homo sapiens (human)			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS		Chou, M.-Y. and Li, H.-C.			
TITLE		Genomic organization and characterization of the human type XX collagen (COL21A1) gene			
JOURNAL		Genomics 79 (3), 395-401 (2002)			
MEDLINE		21853310			
PUBMED		11863369			
REFERENCE		Chou, M.-Y.			
AUTHORS		Chou, M.-Y.			
TITLE		Cloning and identification of a novel human alpha 1 chain-like collagen (COL1A1) gene, a new member of the FACIT family			
JOURNAL		Unpublished			
REFERENCE		3 (bases 1 to 4122)			
AUTHORS		Chou, M.-Y.			
TITLE		Direct Submission			
JOURNAL		Submitted (20-DEC-2000) Biomedical Engineering Center, Industrial Technology Research Institute, 195-85, Sec. 4, Chung Hsing Rd., Chutung, Hsinchu 310, Taiwan			
FEATURES		Location/Qualifiers			
source		1..4122			
		/organism="Homo sapiens"			

/mol_type="mRNA"					
/db_xref="taxon:9606"					
/chromosome="6"					
/map="6p12.3-p11.2"					
1..4122					
/gene="COL1A1"					
163..3327					
/codon_start=1					
/product="alpha 1 chain-like collagen COL1A1 precursor"					
/protein_id="AA150033.1"					
/db_xref="GI:17974510"					
/translation="MAKVTITELCVLILNLSVLAEDGVRSSCRTPATDELIDG					
SYSGPFIKWKVNIITKFDLFAKSRFLTKIAVLTVDGSKQDDVDAQAARQ					
SKITLFAIGVSEEDAEALRAINKPSTVYVYEDYIA-SKIRVNVKLCIESVCP					
TRIPVAARDERGFLLGLDVKVKKRQLSPKKIKGVEYTKVDLSLTSVFPFG					
LPSPVVFSTQRFVKK-LMLWRLITIDGRPQIAVTLNGVDKILLFTTTSVNGSQV					
TFANPVNVTLDDEGWHQ-RLLAVTODVTLVDDQIENKPLHVLGLINGQTIQKY					
SGKETVQFDVQKLRIYCDPQNNREYACETPGFCLNGPSDVSTPAPCIQPKGRL					
QGPKDRPLPQNPYGPQPGQDQKPGYQGIAGTGVFGSGPGIOGARGPGYKGPGRD					
GDKRGRLPFPGLHGMFGSGAGMAGKSGPGFYOKKGAKEKAGNAGPFGPAG					
EPGRHGKDLMSGPGFGEAGSPAGQDGTGEPGIPFGNRMKGQKEIGTPGQ					
BQKGAQPMGLMGSNPSGPGFPGSKSGEFGIQMPGASGLKGPATGSGPSP					
GYMGLPGIKGKKGDKNGEKIGIQKQKNGRQGIQGGQIQGHGAKGREGKEPG					
VWGLGSGESGVDGLMGAPGAPGPGPGPGPGPGPGPGPGPGPGPGPGPGPG					
PAQLPVLQSGRIKNCCHLQSGSGPGIPGPGPGIPGPGPGPGPGPGPGPGPG					
*PQPGVRLGKGLPGNKGKSGSGYSGEQQPGPGPGPGPGPGPGPGPGPGPG					
DDHGKPGIQGPGPGPG:CDPSLCFSVIARNDPFRKGFNY"					
163..228					
/gene="COL1A1"					
4072..4077					
/gene="COL1A1"					
BASE COUNT	1280 a	759 c	996 g	1077 t	
ORIGIN					
Query Match:	100.0%	Score 2865;	DB 9;	Length 4122;	
Best Local Similarity	100.0%	Pred. No. 0;			
Matches 2865;	Conservative	0;	Mismatches	0;	Gaps
0;					
Qy	1	ATGGCTCACTATATACATTTCTCTGTCATGCTTTTGGTGTCTTCTTCAAAATCTGTG	60		
Db	163	ATGGCTCACTAATTACATTTCTCTGTCATGCTTTTGGTGTCTTCTTCAAAATCTGTG	222		
Qy	61	TTAGCTGGAAGTGGGGAAGTAGATCAAGTGTCTGCTACTGCTCGACAGATTTAGTTTC	120		
Db	223	TTAGCTGGAAGTGGGGAAGTAGATCAAGTGTCTGCTACTGCTCGACAGATTTAGTTTC	282		
Qy	121	ATCTTAGATGGCTCTTATAGTGTGGCCGAGAAAACITTTGAATAGTGAANAAGTGGCTT	180		
Db	283	ATCTTAGATGGCTCTTATAGTGTGGCCGAGAAAACITTTGAATAGTGAANAAGTGGCTT	342		
Qy	181	GTCATATCACAAAAACITTTGACATAGGCGCGAAGTTTATCAAGTTGGAGTGTCTCAA	240		
Db	343	GTCATATCACAAAAACITTTGACATAGGCGCGAAGTTTATCAAGTTGGAGTGTCTCAA	402		
Qy	241	TATAGTACTACCTGCTGCTGGAGATTCCTCTCGGAGCTATGATCAGAGAACATTTG	300		
Db	403	TATAGTACTACCTGCTGCTGGAGATTCCTCTCGGAGCTATGATCAGAGAACATTTG	462		
Qy	301	ACGGCAGCAGTGGAACTCATCTCTACTTAGAGAGAACACAAAGACAGAGGAGCCATC	360		
Db	463	ACGGCAGCAGTGGAACTCATCTCTACTTAGAGAGAACACAAAGACAGAGGAGCCATC	522		
Qy	361	CAGTTTGGCTCGATTACCTTTTGGCCAAAGTCTCAGCATTTCTTACTAGATAGCAGTG	420		
Db	523	CAGTTTGGCTCGATTACCTTTTGGCCAAAGTCTCAGCATTTCTTACTAGATAGCAGTG	582		
Qy	421	GTACTTACGATGGCAATCCCAAGATGACCTCAAGATGACGCTCAAGCAGCAAGAT	480		
Db	583	GTACTTACGATGGCAATCCCAAGATGACCTCAAGATGACGCTCAAGCAGCAAGAT	642		
Qy	481	AGTAAGATAACATTTATTTGCTTGTGTTTGGTTTTCAGAAAACAGAGATGCCAACTTAGA	540		

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: October 2, 2003, 19:29:48 ; Search time 10:31 Seconds
(without alignments)
11569.053 Million cell updates/sec
Title: US-09-996-611B-5
Perfect score: 2865
Sequence: 1 atggctcattatattacatt.....gaaaggacaaactattag 2865
Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0
Searched: 2888711 seqs, 2045481386 residues
Word size : 0
Total number of hits satisfying chosen parameters: 5777422
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sv:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2865	100.0	4122	9	AF310693	AF310693 Homo sapi
2	2112	73.7	4160	9	AF414088	AF414088 Homo sapi
3	2061	71.9	4160	6	AX086127	AX086127 Sequence
4	2061	71.9	4160	9	HS086157	AX136624 Homo sapi
5	974	34.0	1184	6	AX430410	AX430410 Sequence
6	923	32.2	1169	6	AX430281	AX430281 Sequence
7	890	30.7	2295	9	AK096444	AK096444 Homo sapi
8	749	26.1	1115	6	AX522338	AX522338 Sequence
9	552	19.3	11000	9	AF439327_2	Continuation (3 of
10	552	19.3	126501	9	HS708F5	AX721209 Sequence
11	382	13.3	1414	6	AX721209	AX721209 Human DNA
12	252	9.8	447	6	BDC58397	BDC58397 Secrete
13	204	7.1	81696	9	AF438327_3	Continuation (4 of
14	204	7.1	118431	9	HS682115	ALC34452 Human DNA
15	131	4.6	246	6	B0053596	B0053596 Sequence
16	63	2.2	1101	6	AX721212	AX721212 Sequence
17	58	2.0	25483	9	AL450431	AL450431 Human DNA
18	58	2.0	50612	9	AL513530	AL513530 Human DNA
19	24	0.8	128960	9	HSJ919F19	AL109947 Human DNA
20	23	0.8	67884	2	AC090561	ACC090561 Homo sapi
21	23	0.8	132870	8	AC119149	AC119149 Oryza sat
22	23	0.8	15880	9	AF005716	AF005716 Homo sapi
23	23	0.8	166093	9	AC106013	AC106013 Homo sapi
24	23	0.8	188326	9	AC104958	AC104958 Homo sapi
25	23	0.8	304110	8	AE017074	AE017074 Oryza sat
26	22	0.8	2733	10	AY158896	AY158896 Rattus no
27	22	0.8	5551	10	AF272662	AF272662 Rattus no
28	22	0.8	6114	10	CRU8A1V	M76730 Chinese ham
29	22	0.8	7228	10	AF049880	AF049880 Mus muscu
30	22	0.8	51870	2	AC104220	AC104220 Mus muscu
31	22	0.8	52855	2	AC129491	AC129491 Mus muscu
32	22	0.8	111799	9	HSJ850015	AL049637 Human DNA
33	22	0.8	127277	9	AC010175	AC010175 Homo sapi
34	22	0.8	139273	9	AL596275	AL596275 Human DNA
35	22	0.8	140119	10	AL928839	AL928839 Yousc DNA
36	22	0.8	15807	10	AC122459	AC122459 Mus muscu
37	22	0.8	15793	2	AC102109	AC102109 Mus muscu
38	22	0.8	162572	9	AC009533	AC009533 Homo sapi
39	22	0.8	166751	10	AC087066	AC087066 Rattus no
40	22	0.8	169947	2	AC102838	AC102838 Mus muscu
41	22	0.8	176694	2	AC053494	AC053494 Homo sapi
42	22	0.8	177135	10	AL663099	AL663099 Mouse DNA
43	22	0.8	181742	10	AC122397	AC122397 Mus muscu
44	22	0.8	187999	2	AC026928	AC026928 Homo sapi
45	22	0.8	188162	9	GNS05TDR	AL357092 Human chr
46	22	0.8	194477	2	BX248318	BX248318 Danio rer
47	22	0.8	206032	2	AC134799	AC134799 Rattus no
48	22	0.8	207366	10	AL606750	AL606750 Mouse DNA
49	22	0.8	208122	2	AC121207	AC121207 Rattus no
50	22	0.8	210585	2	AC107788	AC107788 Mus muscu
51	22	0.8	219899	10	AL663039	AL663039 Mouse DNA
52	22	0.8	232649	2	AC108267	AC108267 Rattus no
53	22	0.8	241372	2	AC114198	AC114198 Rattus no
54	22	0.8	245306	2	AC123138	AC123138 Rattus no
55	22	0.8	250023	2	AC111499	AC111499 Rattus no
56	22	0.8	254778	2	AC126532	AC126532 Rattus no
57	22	0.8	255369	2	AC095247	AC095247 Rattus no
58	22	0.8	263088	2	AC130872	AC130872 Rattus no
59	21	0.7	6273	10	SC052161	SC052161 Mus muscu
60	21	0.7	7057	10	MUSA1CC	D38162 Mouse mRNA
61	21	0.7	42613	3	AF106575	AF106575 Caenorhab
62	21	0.7	82700	2	AC015635	AC015635 Homo sapi
63	21	0.7	83000	9	AP033055	AP033055 Homo sapi
64	21	0.7	110200	2	AC114905_2	Continuation (3 of
65	21	0.7	119265	9	AL158046	AL158046 Human DNA

```

Qy 2338 AAGCCCGAAGAGAGTTTTCAGAACAAATTTATTCGACAAAGTTTGCACAGATGTAATAAGA 2397
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2409 CCTCCGGCAGAGCGCGACCGACCAATCAAGCAGGTTTGCATGAGAGTCGTGCAA 2468
Qy 2398 GCCCAGCTACCAAGTCTTACTTCAGAGTGGAAATTAGAAATTTGTGATCATTTGCCCTGTCC 2457
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2469 GAGCATTTTGGGGAATGGCAGCCAGCTCAAGAGACCAGACACAGAGCCCT----- 2522
Qy 2458 CAACATGGCTCCCGGGTATTCCTGGGCCACCTGGTCCGATAGGCCCCAGAGGGTCCCAGA 2517
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2521 ----CTGGTCTTCTCTGGAGGCTTGGCCGCCCTGGGCTCCGGGCCCTCGAGAGAA 2576
Qy 2518 GGATTACCTGTTTCCAGGAGAGATGGTGTTCCTGGATTAGTGGGTGTCCCTGACGT 2577
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2577 GGTTCCTGCTCAGATGGGAATCCGTGGTCTCCAGGCATTAAAGGGTCCCTGGTCT 2636
Qy 2578 CCAGGTGTCAGAGATTAAAGGCTTACCAGGAAGAAATGGGAAAAGGGAGCCCAAGGG 2637
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2637 CTTGGCTTAAGAGGACCTTAAGGAGACTTGGGAGAAAAGGAGACGTGTCCTCCAGGA 2696
Qy 2638 TTTGGGTATCCTGGAGAACAAAGTCTCTGTGTCCCGAGTCCAGAGGCCCTCTCTGA 2697
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2697 AGAGTCTTAAGGTTTTCCTGGAGGTAGGTCTCCAGGTGACCCAGGCCCTGCCAGC 2756
Qy 2698 ATAAGCAAAGAGTCTCCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGACCATGGA 2757
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2757 TATGGGAAAATGGCCGTGACGAGAGCAAGGTCCCGGGAGTGGCAGGAATTCCTGGT 2816
Qy 2758 AAACCTGGAATCCCAAGGGCAACAGGCCCCCGAGGCATCTGGACCCATC 2807
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2817 GTGCTGGACCCCGAGTCTCCAGGCCCTCTGGGTTCGTGAGCAAGC 2866

```

Search completed: October 2, 2003, 23:48:11
Job time : 5323 secs

Db 972 GAACAGGTTCTGAGGAGACAGAGGCAATTCAGGTTCCGAGGAATTCCTGGATCCCA 1031
Qy 1678 GGCCTCCCTGACCTGCTGGAGAACACAGGAGACATCGAAAGGATGATTAATGGGTAGT 1737
Db 1032 GGGCAAAAGAGAGACAGGCTTGCACAGGTGTAGATGGCCGAGATGGAATACAGGAATG 1091
Qy 1738 CCGCGTTTCAAGGAGAGAGAGATCCCTGCTGCTCCCGGGCAGGATGGACACGGGA 1797
Db 1092 CCGCGGACAAAGAGGTGAAGCAGGGAAGCCTGGACCCCGGGACCTGGGATTCGAGGCG 1151
Qy 1798 GAGCTCGAATCCCAAGATTTCTCTGGAAACCGAGGATTAATGGGCCCAAAAGGAGAGAAAT 1857
Db 1152 TTACAGGTGTCTCTGATCCCTGCTGTCGAAAGGTCTTCCCGTGAAGAGGTAAACG 1211
Qy 1858 GGGCTCTCAGACAGCAAGCAAGAAAGAGAGCCCGAGGATGCTGTTTAATGGGAAGC 1917
Db 1212 GGTCTCTCAGGAAAGCCCGGTCACTTGGGAATTCAGGGAACACAGGCCAACAGAGGCC 1271
Qy 1918 AATGGCTCACAGGCCAGCTCGGAACACCGGGATCTAAGGGAAGCAAGTGGAACCTGA 1977
Db 1272 CCAGAGAGGTTGAACCTCGGGACCCAGGGCTTCCAGGCAATAGAGCCCGGTAGA 1331
Qy 1978 ATTCAAGGATGCTCGGGCTTCTGGGCTCAAGGAGAACCCAGGCAACAGGTTCCCA 2037
Db 1332 CCAGAGGCTTCCAGGCATACCAAGGGAATCTGTTGGCAGCCCTGGCTTCTCT 1391
Qy 2038 GGAGAACCAAGATACATGGGTTTACCGCGGATTCAGGGAAGGAGGAGCAAGGAAT 2097
Db 1392 GGTCTGCTGGGCTTCTGAGTCTCTGGAATGAAGGAGACAGGCTGATTTGGTGA 1451
Qy 2098 CAAGGTGAAGAGTATTCAGGTCGAAAGGAGAAATGAAGAGACAGGGAATTCAGGG 2157
Db 1452 CCGGCTCCAGGCTGAACAGGCTGCTCTGTTGAAGAGGTGAAGCAGGCAAGGGCT 1511
Qy 2158 CAACAGGGAATTCAGGCTCATGCTGCAAGAGGAGAGAGGTGAAGGAGGACCT 2217
Db 1512 GACCTTGGAGATATGGACACCTTGGCCCAAGAGGATCTGTGGTAACTCCCGGGAGCCG 1571
Qy 2218 GGTCTCGAGTGCCTTGGATCAAGAGGAGATCTGGGTGGATGGCTTGTATGGGGCC 2277
Db 1572 GGTCTGAGGGGCTGAAGGAATCAGAGGCTTCTGGAGTGAAGGACCAAGAGGACCA 1631
Qy 2278 GCAGTCTTAAGGGAACCTCGGGATCTCAGGCTCCTCAGGACCCCGAGTTGGATGG 2337
Db 1632 CTGAGCCCGGCGATCGAGGAGATCAGGCTGCACTGGCTGGCTGATTCAGGAGC 1691
Qy 2338 AAGCCCGAAGAGATTTTCAGAACAAATTTATTCGACAGTTTGCACAGATGTAATAGA 2397
Db 1692 CCTCGGCGAGGCGCCGACCGACGACATCAAGCAGGTTTGCATGAGAGTCGTGCA 1751
Qy 2398 GCCAGCTACAGCTTACTTCAGAGTGGAGAAATTAAGAAATTTGATCATTCCTGCTC 2457
Db 1752 GAGCATTTTGGGAATATGGCAGGAGCTCAAGAGACACAGAGAGCCT----- 1803
Qy 2458 CAACATGCTCCCGGTTATTCCTGGGCACTCTGGTCCGATAGCCCGAGAGGTCACGA 2517
Db 1804 ----CTGGTCTCTCGAGGCTTGGCCCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 1859
Qy 2518 GGATTAATCTGTTTCCAGGAGAGATGGTGTCTCTGGATAGTGGGTGCTCTGAGCT 2577
Db 1860 GGTTCCTCTGATGGGAATCCGTGCTCTCCAGGCATTAAGGCTCCCTCTGCTGCT 1919
Qy 2578 CCAGTGTACAGGATTAAGGCTTACAGGAGAAATGGGAAAGAGGAGCCAGG 2637
Db 1920 CTGCTTAAGAGACCTTAAGGAGACTTGGGAGAAAGAGAACTGGTCTCTCAGGA 1979
Qy 2638 TTGAGTATCTCGAGAACAAAGGCTCTCTGCTCCCGAGGTCCAGAGGCTCTCTGGA 2697
Db 1980 AGAGTCTTAAGGTTTCTGCTGGAGCGATAGTCTCTCCAGGTGACCGGCTCTGCGAG 2039
Qy 2698 ATAAGCAAGAGGCTCTCAGGAGACCCAGGCTCTCTGCGAAGATGAGACACATGA 2757
Db 2040 TATGGAAAAATGGCGTGAAGGAGAGCAAGGCTCCCGGAGTGGCAGGAATTCCTGCT 2099

Qy 2758 AAACCTGGAAATCCAAAGGCAACCCAGGCCCCAGGAGCATCTGCGACCCATC 2807
Db 2100 GTGCTGGACCCCGAGGCTCTCAGGCCCTCTGCGATCTGTSAGCCAGC 2149

RESULT 15
AK031071 3405 bp mRNA linear HTC 05-DEC-2002
LOCUS Mus musculus 13 days embryo forelimb cDNA, RIKEN full-length
DEFINITION enriched library, clone:5930402D07 product:procol-agen, type IX,
alpha 1, full insert sequence.

AK031071.1 GI:26326988
VERSION H:C: CAP trapper.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteiostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Carninci, P., Shibata, Y., Hayata, N., Sugahara, Y., Shibata, K.,
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999).
99279253
10349636

REFERENCE
2
10349636

REFERENCE
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishire, F., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Chara, E., Wachihi, M.,
Yoneda, Y., Shikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-184-format
Sequencing pipeline with 384 multiplexed capillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

REFERENCE
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Azawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yananaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, K.,
Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, T., Pesole, G.,
Quackenbush, J., Schram, L. M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furumori, M., Aono, H.,
Bardarelli, R., Barsh, O., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, K. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kotsuki, S.
and Hayashizaki, Y.

REFERENCE
5
Function: annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851

REFERENCE
5
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.

```

Db 1246 CCAGGAGAGGTTGGACCTCGGGACCCAGGGGCTTCGAGCAGTAGAGGCCGGTAGGA 1305
Qy 1978 ATTCAAGGAGATGCTGGGGCTTCTGGGCTCAAGGGAGAACCCAGGAGCAACGGTTCCCA 2037
Db 1306 CCAGAGAGGCTCTCCAGGCATACCCAGGAACTGGATCTGTTGGCAGCCCTGGCTTCT 1365
Qy 2038 GGAGACCAAGGATACATGGTTTACCGGGATTCAGGAAAGAGGGGCAAGAGAAAT 2037
Db 1366 GGCTTGGCTGGGCCCCCTGGACTTCTGGAAATGAAGGAGACAGGGGTGATTTGGTGA 1425
Qy 2098 CAAAGTGAAAGAGTATTCAAGGCTCAAGAGGAGAAATGGAAGACAGGAAATCCAGGG 2157
Db 1426 CCGGTCCTCAAGGGTGAACAGGGTGCCTCTGGTGAAGAGTGAAGCAGGAGCAAGGGT 1485
Qy 2158 CAACAGGAAATCAAGGCATCATGTGCAAGAGGAGAGAGAGGTGAAGAGGAGAACCT 2217
Db 1486 GACCTTGGAGATATGGCAACCTGGCCCCAAGAGGATCTGGGTAAACCCCGGGAGCG 1545
Qy 2218 GGTGTCGAGGTGCCATTGGATCAAGAGGAGAAATCTGGGGTGGATGGCTTGATGGGGCC 2277
Db 1546 GGTCTGAGGGGCTGAAGGATCAGAGGCTTCTGGAGTGAAGGACCAAGAGACCA 1605
Qy 2278 GCAAGTCTAAGGGGCAACCTGGGGATCCAGGTCTCAGGACCCCGAGGTTTGGATGG 2337
Db 1606 CTTGGACCCCGGGGATGAGGGAGATCAGGGTGCCTGGCTGCTGGTATCCAGGGC 1665
Qy 2338 AAGCCGGAAGAGATTTTCAGAACAAATTTTCGACAAAGTTTGCACAGATGTAATAGA 2397
Db 1666 CTTGGGAGAGCGGGACCGACAGCATCAAGCAGTTTGCATGAGAGTGTGGA 1725
Qy 2398 GCCAGTACCAAGTCTTACTTTCAGAGTGAAGAAATAGAAATTTGATCATTCCTGTCC 2457
Db 1726 GAGCAATTTGCGAAATGGCAGCAGCTCAAGAGACCCAGACACAGAGGCTT----- 1777
Qy 2458 CAACATGGCTCCCGGGTATTTCTGGCCACCTGTCCGATAGGCCAGAGGTCCTCAGA 2517
Db 1778 -----CTGGTCTTCTGGAGGCTTGGCCCCCTGGGCTTCCGGGCCCTCCCTGGAGAA 1833
Qy 2518 GATATACCTGTTTGCAGAGAGATGTTGTTCTTGGATTTAGTGGTGTCTCTGGAGCT 2577
Db 1834 GGTTCCTGTCAGATGGAAATCGGTGTTCTCCAGGCAATTAAGGGTCCCTGCTGCT 1893
Qy 2578 CCAGTCTCAGAGATTAAGGCTTACCAGGAGAAATGGGGAAGAGGAGGAGGAGG 2637
Db 1894 CTTGGCTTAAGAGACCTTAAGGAGACTTGGGAGAAAGAGGAGAGTGTCTCTCAGA 1953
Qy 2638 TTTGGGTATCTTGAGAACAGGTCTCTCTGTTCTCCCGAGTCCAGAGGGCCCTCTGGA 2697
Db 1954 AGAGTCTTAAGGTTTGGCTTGGAGGATAGGTCTCCAGGTGAACCCAGGCTCCAGC 2013
Qy 2698 ATAAGCAAGAGTCTCCAGGAGCCAGGTCTCTCTGGCAAGATGGAGACCATGGA 2757
Db 2014 TATGGGAAATGGCGTGAAGGAGAGAGGTCTCCCGGGAGTGGAGAAATCTCTGT 2073
Qy 2758 AAACCTGGAATCCAAGGCAACAGGCCCCCAGGCAATCTGGAGCCCATC 2807
Db 2074 GTGCTGAGCCCCAGGTCTCCAGGCTCTCTGAGGCTTCTGTGAGCAGC 2123

RESULT 14
BC038017 3035 bp mRNA linear HTC 23-SEP-2002
LOCUS Mus musculus, similar to procollagen, type IX, alpha 1, clone
DEFINITION IMAGE:3592397, mRNA.
ACCESSION BC038017
VERSION BC038017.1 GI:23274264
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3035)
AUTHORS Strausberg,R.

```

TITLE
JOURNAL

REMARK
COMMENT:

Direct Submission
Submitted (20-SEP-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland
Web site: <http://www.nisc.nih.gov/>
Contact: misc_mgc@nigri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dierich,N.L., Grantham,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Xastrian,S.D., McCloskey,J.C.,
McDowell,C., Pearson,R., Sraatropop,S., Thomas,P.J., Touchman,J.W.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,S.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 80 Row: c Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein

This clone has the following problem: frame shifted.

FEATURES
source
..3035
Location/Qualifiers

```

0 * /organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3592397"
/tissue_type="Mammary tumor. Metathelionien-TGF alpha
node1. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"

```

BASE COUNT 767 a 716 c 915 g 637 t
ORIGIN

```

Query Match 10.2%; Score 291.2; DB 11; Length 3035;
Best Local Similarity 50.3%; Pred. No. 1e-44;
Matches 750; Conservative C; Mismatches 728; Indels 12; Gaps 1;

Qy 1318 CCTGTATTGTTCTCCCGGAAACAGGACTTCAAGGCCCAAGGTGACCTGTA 1377
Db 672 CCCAATTCCTGCCACCAAGGTGCTCTGGATATCAGGCTTACAGGCTCAGGCGGCAT 731
Qy 1378 CTGGGAAACCTGGCTACCTGGCAACCTGTCAGATGTAAGCTGGATATCAGGGA 1437
Db 732 AAAGGGCCGAGAGAAATCGGCGAGCCAGGAGACAGGACACAGGAGGAGAGG 791
Qy 1438 ATTGCAGGAGACACAGGTGTTCCAGGATCTCCAGGATACAGGAGCTCAGGACTACCA 1497
Db 792 GACCAAGGGGAACTGGGAGAAAGTTGGCGATCAAGGACCTCAGGACCTCAGGCTTGA 851
Qy 1498 GGTTCACAAAGGAGAACACAGGCGAGATGGTGACAGGCTGATCTGGACTTCCTGTTT 1557
Db 852 GGCATCTCTGGCATAGTTGGAGACAAAGGAGAAAGGCTGCTCGGGATTTGATGGAG 911
Qy 1558 CTTGGCTTTCATGGCATGCCAGGATCAAGGGTGAATGGGTGCCAAGAGAGACAAGGA 1617
Db 912 CTTGGACCTCAGGGCATTCAGGTGTCAGCTGGTATCAAGGACAGGAGGAGGCTCCAGGA 971
Qy 1618 TCACCTGGATTTTATGGCAAAAGGTCGCAAGGTCGAAAGGAGGATGCTGGCTTCCT 1677

```


Db 651 CCCAATTCCTGCCACCCAGTCTCGTCTGGATATCCAGGCTACCAAGGCATCAGGGGCAT 710
 QY 1378 CTTGGGACCCCTGGCTACCTCGGACAACTGTGTCAAGATGTGTAAAGCTGGATATCAGGA 1437
 Db 711 AAAGGGCGGAAGGAGAAATCCGCGAGCCAGGACCAAGGACCAAGGGGTGAAGAGGT 770
 QY 1438 ATTGACGACACACAGGTGTTCCAGGATCTCCAGGAATACAAAGGAGCTCCAGGACTACCA 1497
 Db 771 GACCAGGGGAACTGGGAAAGTTGGCGATCAAGGACCTCCAGGACCTCAGGTCTGAGA 830
 QY 1498 GGTTCAAAGGAGAACAGGGCGAGATGTGTGACAAAGGTGTATGTGTGACTTCTGTGTTT 1557
 Db 831 GGCATCACTGGCATAGTTGGAGACAAAGGAGAAAGGGTGTCTGGGGATTTGATGGAGAG 890
 QY 1558 CTTGGGCTTCATGGGATCCAGGATCAAGGGGTGAATGGGTGCCAAGGAGACAAAGGA 1617
 Db 891 CTTGGACCTCAGGGCATTCAGGTGCGCTGTGTATCAAGGACACGAGGCGCCCTCCAGGA 950
 QY 1618 TCACCTGGATTTTATGGCAAAAGGGTGCAAAAGGTGAAAAGGGGAATGCTGGCTTCCCT 1677
 Db 951 GAAACAGGTCCTGAGGAGACAGAGCATTTCAAGGTTCAGGAGAAATTCCTGGATCCCA 1010
 QY 1678 GGCCTCCCTGGACCTGCTGGAGAACAGGAGACATGGAAGGATGGATTAATGGGTAGT 1737
 Db 1011 GGGCCAAAAGGAGACACGGGCTTGGCAGGTGTAGATGGCCGAGATGGAATACCAAGGAATG 1070
 QY 1738 CCGGTTTCAAGGGAGACAGGATCCCTGCTGCTCGGGGCGAGATGGAAACACGGGA 1797
 Db 1071 CCGGGACAAAGGGTGAACAGGAGAGCTGACCCCGGGAGCTGGGATTCGAGGGC 1130
 QY 1798 GAGCTGGAAATCCAGGATTTCTGTGAAACCCAGGATTAATGGGCCAAAGGGAGAAAT 1857
 Db 1131 TTACCAAGGTGCTCCCTGGAATCCTGTGCAAAAGGTGTGCGGTGAAAAGGGTAACACG 1190
 QY 1858 GGGCTCCAGGACAGCAAGGAGAAAGGAGCCAGGGATGCTGCTGGTTTATGGGAGC 1917
 Db 1191 GGTGCTCCAGGAAGCCCGTCACTGTTGGGAAGTTCAGGGAACACAGGCCCAACAGGGCCC 1250
 QY 1918 AATGGCTCACAGGGCAGCTCGAACACCGGGATCTAAGGGAAGCAAGGTGAACCTGGA 1977
 Db 1251 CCAGGAGAGGTTGGACTTCGGGACCCAGGGGCTTCCAGGCAGTAGAGGCCCGGTAGGA 1310
 QY 1978 ATTCAAGGATGCTGGGCTTCTGGCTCAAGGAGAACACAGGAGAACCGGTTCCCA 2037
 Db 1311 CCAGAAGGGTCTCCAGGCATACAGGGAACACTGGGATCTGTGGCAGCCCTGGCTTCTCT 1370
 QY 2038 GGAGAACACAGGATACATGGGTTTACCCGGATTTCAGGGAAGGAGGAGCAAGGAAAT 2097
 Db 1371 GGTGCTGCTGGGCTTCTGGACTTCTGGATGAAGGAGACAGGGGTGTATTTGGTGAA 1430
 QY 2098 CAAGGTGAAGAGTATTCAAGGTCMAAAGGAGAAATGGAAGACAGGGAATCCAGGG 2157
 Db 1431 CCGGTCCTCAAGGGTGAACAGGGTGGCTCTGTTGAAGAGGTGAAGAGGAGCAGAGGGGT 1490
 QY 2158 CAACAGGGAATCAAGGCATCATGTGTGCAAAAGGAGAGAGAGGTGAAGAGGAGAACT 2217
 Db 1491 GACTTGGAGATATGGGACAACTGGCCCAAGAGGATCTGTGGGTAAACCCGGGAGCG 1550
 QY 2218 GGTGTCCGAGGTGCCATTTGGATCAAAAGGAGAAATCTGGGGTGGATGGGTGTATGGGGCCC 2277
 Db 1551 GGTCTGAGGGGGCTCAAGGATCAAGGGCTTCTGGAGTGAAGGACCAAGAGGACCA 1610
 QY 2278 GCAAGTCTTAAGGGGCAACTGGGATCCAGCTCTCAGGACCCCGAGTTTGGATGGG 2337
 Db 1611 CTTGGACCCCGGGGATGAGGAGATCAGGTCGACCTGGCTCTGGTATCCAGGGC 1670
 QY 2338 AAGCCCGGAGAGAGTTCAGAACAAATTTATTCGACAGTTTGCACAGATGTAATAGA 2397
 Db 1671 CTTCCGGGACAGCGCGGACCCAGCAGCATCAAGCAGGTTTGCATGAGAGTCTGTGCA 1730
 QY 2398 GCCCAGCTACCACTTACTTTCAGATGGAAGAAATTTAGAAATTTGTATCATTTGCTGTCC 2457

Db 1731 GAGCATTTTTCGGAATGGCAGCAGCCTCAAGAGACCAAGACACAGGAGCCT----- 1782
 QY 2458 CAACATGGCTCCCGGGTATTCTCTGGGCGACCTGGTCGATAGGCCAGAGGGTCCCGAGA 2517
 Db 1783 ----CTGGTCTTCTGGGAGGCTGGCCCCCTCGGGCTCGGGCCCTCCCTGGAGAGAT 1838
 QY 2518 GGATTAACCTGGTTTCCAGGAAGAGATGGTGTCTCTGGATAGTGCGTGTCTCTGGACGT 2577
 Db 1839 GGTTCCTCTGTTCAGATGGGAATCGTGTCTCTCCAGGCAATTAAGGTCCTCCCTGGTGT 1898
 QY 2578 CAGGTGTTCAGAGGATTTAAAGGCTTACAGGAAGAAATGSGAAAAGGAGGCAAGGG 2637
 Db 1899 CTTGGCTTAAGAGGACCTTAAGGAGACTTGGGAGAAAAGGAGAACTGGTCTCCAGGA 1958
 QY 2638 TTTGGGTATCTCGGAGAACAAAGGTCTCTGCTGCCCGCAGGTCCAGAGGGGCCCTCTGGA 2697
 Db 1959 AGAGTCTTAAGGTTTGCCTGGAGCGATAGTCTCCAGGTGACCCAGGCCCTCGCAGC 2018
 QY 2698 ATAAGCAAGAGGTCTCTCAGGAGACCCAGGTCTCTCGCAAGATGAGACCATGGA 2757
 Db 2019 TATGGAAAATGCGCGTACGAGCAAGTCCCGCGAGTGGCAGGAATCTCTGT 2078
 QY 2758 AAACCTGGATCCAGGGCAACAGGCCCGCCAGGCATCTCGACCCATC 2807
 Db 2078 GTGCCGTGACCCCAAGTCTCTCCAGGCCCTCTGGTCTGTGAGCCAGC 2128

RESULT 13

AK011547 2956 bp rRNA linear HTC 07-DEC-2002
 LOCUS Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
 DEFINITION enriched library, clone:2610025B01 product:procollagen, type IX,
 alpha 1, full insert sequence.

ACCESSION

AK011547
 VERSION AK011547.2 GI:26374667
 KEYWORDS HTC; CAP trapper.

SOURCE

Mus musculus
 ORGANISM Mus musculus (house mouse)

REFERENCE

1 Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 2
 REFERENCE
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

20499374
 MEDLINE 11042159
 PUBMED

REFERENCE

3
 REFERENCE
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multiplexed sequencer
 Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

20530913
 MEDLINE
 PUBMED :076861

REFERENCE

4
 REFERENCE
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, S., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamahata, I.,
 Saito, T., Okazaki, Y., Gojcbori, T., Bono, H., Kasukawa, T., Saito, R.,
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,

Db 366 GTCTGACTGGCATCCCGGGTCGCCCTGGGGCTCGTGGGACAGAGGANTGCCCTGGAAAA 425

QY 2615 ATGGGAAAAGGAGGCAAGGGTTTGGGTATCTCTGGAGAACAGAGTCCCTCGGTCCCC 2674

Db 426 ATGGTGAAAGGCAATCAAGGAATTTGGGTCTCTGGGATCCAGGCCCCACGAGACCGC 485

QY 2675 CAGGTCAGAGGGCCCTCTCTGGAAATAGCAAGAGAGTCTCCAGGAGACCCAGGTCTCC 2734

Db 486 CAGGTCCTGAAGGTCCCGCGATAGTAAGAAAGGACGCCCTGGAGAGCGTGGGCGC 545

QY 2735 CTGCAAGATGGAGCATCGRAAACCTGATCCAGGCAACACAGGCCCCCGGCA 2794

Db 546 CTGTAAGATGGAGATCGTGGCAGTCCCGGAATGCCAGACACGAGTCCGACCCCTGGAA 605

QY 2795 TCTGCGACCCATCATCTATGTTTTAGTGTAAATTCGCAAGAGATCCGTTTCAGAAAAAGGAC 2854

Db 606 TCTGTGCCCATCATCTGTTTCAGTGTGTCATTGTCGGAAGAGATCCATTAGAAAGGAC 665

QY 2855 CAAA 2858

Db 666 CAAA 669

RESULT 9
BG699698
LOCUS
DEFINITION 602681559f1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:481419 5',
mRNA sequence.
ACCESSION BG699698
VERSION BG699698.1 GI:13968272
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 744)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM10709 row: e column: 16
High quality sequence stop: 596.
Location/Qualifiers

FEATURES
source
1..744
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:481419"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcagag
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
202 a 206 c 177 g 159 t

Query Match 12.0%; Score 344; DB 10; Length 744;

Best Local Similarity 96.4%; Pred. No. 9.2e-55;
Matches 374; Conservative 0; Mismatches 10; Indels 4; Gaps 2;

QY 1 ATGGCTCACTATATTACATTTCTCTGCATGGTTTTGGTCTCTTCCTCAGAAATCTCTG 60

Db 193 ATGGCTCACTATATTACATTTCTCTGCATGGTTTTGGTCTCTTCCTCAGAAATCTCTG 252

QY 61 TTAGCTGAAGATGGGGAAGTAAAGTCAAGTTGTCTACTGTCTCCGACAGATTTAGTTTC 120

Db 253 TTAGCTGAAGATGGGGAAGTAAAGTCAAGTTGTCTACTGTCTCCGACAGATTTAGTTTC 312

QY 121 ATCTTAGATGGCTCTTTAAGTGTGGCCGAGAAAACCTTTGAA---ATAGTGAAGAGTGG 177

Db 313 ATCTTAGATGGCTCTTTAAGTGTGGCCGAGAAAACCTTTGAACTAGTGAAGACAGTGG 372

QY 178 CTGTGCAATATCACAAAAAATTTGCACATAGGGCCGAAAGTTTATTCAGATTGGAGTGGTT 237

Db 373 CTGTGCAATATCACAAAAAATTTGCACATAGGGCCGAAAGTTTATTCAGATTGGAGTGGTT 432

QY 238 CAATATAGTACTACCTGTGCTGGAGATTCCTCTCGAAGCTATGATTAGGAGAACAT 297

Db 433 CAATATAGTACTACCTGTGCTGGAGATTCCTCTCGAAGCTATGATTAGGAGAACAT 492

QY 298 TTGACGG-CAGCAGTGGATCCATCTCTACTTACGAGCAACACAGACAGGGAAGGC 356

Db 493 TTGACGGCAGCAGTGGATCCATCTCTACTTACGAGCAACACAGACAGGGAAGGC 552

QY 357 CATCCAGTTTGGCTCGATTACCTTTT 384

Db 553 CATCCAGTTTGGCTCGCATTTACCTTTT 580

RESULT 10
BU108124
LOCUS
DEFINITION 6031112461F1 CSEQCHL12 Gallus gallus cDNA clone ChEST58j15 5', mRNA
sequence.
ACCESSION BU108124
VERSION BU108124.1 GI:25310967
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 760)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.C.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST
)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers

FEATURES
source
1..760
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="ChEST58j15"
/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSEQCHL12"
/note="Organ: heart; Vector: pBluescript II KS(+); Site 1:
ECORI; Site 2: NotI; Modification of pBluescript II KS(+)
[Stratagene] vector to accommodate cDNA produced with the

Seq primer: M13RPI reverse primer (ABI).

```

FEATURES
  source
    1..527
      /location="Qualifiers"
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="cs68a05"
      /tissue_type="RPE/choroid"
      /dev_stage="Adult"
      /lab_host="ENDH10B"
      /clone_lib="Human Retinal pigment epithelium/choroid cDNA
      (Un-normalized, unamplified): cs"
      /note="Organ: Eye; Vector: pCMVSPORT6; Two different donor
      eyes (75-80 years old) yielded approximately 600 mg of
      dissected RPE/choroid tissue. This in turn yielded 340 ug
      of total RNA and 7 ug of mRNA. A directionally cloned cDNA
      library in the pCMVSPORT6 vector was constructed at Life
      Technologies (Rockville, MD; now part of Invitrogen Corp),
      essentially following the protocols of the SuperScript
      Plasmid System (Invitrogen Corp).
      <http://www.invitrogen.com/>. The library code
      designation was cs. For this library, cDNA inserts were
      cloned into the NotI/MluI sites of the vector. EST
      analysis was performed on the unamplified library at the
      NIH Intramural Sequencing Center (NISC)."
BASE COUNT      145 a 118 c 126 g 138 t
ORIGIN
Query Match      13.2% Score 377; DB 14; Length 527;
Best Local Similarity 100.0%; Pred. No. 5,1e-61;
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTCACTATATTACATTTCTCTGATGGTTTTGGTGTCTTCTTCAGAAATCTGTG 50
DB 151 ATGGCTCACTATATTACATTTCTCTGATGGTTTTGGTGTCTTCTTCAGAAATCTGTG 210
QY 61 TTAGCTGAAGTGGGAGTGAAGTCAAGTTGCTGCTACTGCTCCGACAGATTAGTTTTC 120
DB 211 TTAGCTGAAGTGGGAGTGAAGTCAAGTTGCTGCTACTGCTCCGACAGATTAGTTTTC 270
QY 121 ATCTAGATGGCTCTTATAGTGTGGCCAGAAACCTTTGAAATAGTGAAGTGGCTT 180
DB 271 ATCTAGATGGCTCTTATAGTGTGGCCAGAAACCTTTGAAATAGTGAAGTGGCTT 330
QY 181 GTCATATCAAAAACCTTTGACATAGGCCGCAAGTTTATTCAGTTGGTGTCAA 240
DB 331 GTCATATCAAAAACCTTTGACATAGGCCGCAAGTTTATTCAGTTGGTGTCAA 390
QY 241 TATAGTCACTACCTGTGCTGGAGATTCCTCTCGAAGCTATGATTCAGGAGAACATT 300
DB 391 TATAGTCACTACCTGTGCTGGAGATTCCTCTCGAAGCTATGATTCAGGAGAACATT 450
QY 301 ACGGACAGTGGAAATCCATCTCTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 360
DB 451 ACGGACAGTGGAAATCCATCTCTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 510
QY 361 CAGTTTCGGCTCGATTA 377
DB 511 CAGTTTCGGCTCGATTA 527

RESULT 8
LOCUS BU250509
DEFINITION 603404356f1 CSEQCHN35 Gallus gallus cDNA clone CHEST304013 5', mRNA
sequence.
ACCESSION BU250509
VERSION BU250509.1 GI:25501587
KEYWORDS EST
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

```

```

Phasianinae; Gallus.
X (bases 1 to 670)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.C.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
MEDLINE
PUBMED
12445392
COMMENT
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
  1..670
    /organism="Gallus gallus"
    /mol_type="mRNA"
    /strain="White Leghorn, HiseX"
    /db_xref="taxon:9031"
    /clone="CHEST304013"
    /dev_stage="36"
    /lab_host="DH10B"
    /clone_lib="CSEQCHN35"
    /note="Organ: trunk; Vector: pBluescript II KS(+);
    Site 1: EcoRI; Site 2: NotI; This normalized library was
    constructed from 1 million independent clones. cDNA
    synthesis was initiated using an oligo(dT) primer, using
    methylated C in the first strand synthesis reaction.
    Following this first strand reaction, double-stranded cDNA
    was blunted, ligated to NotI adapters, digested with EcoRI
    , size-selected, and cloned into the NotI and EcoRI
    compatible sites of a custom modified MCS of the
    pBluescript (KS+) vector. The library was normalized in 2
    rounds using conditions adapted from Soares et al., PNAS
    (1994) 91: 9228-9232 and Donaldo et al., Genome Research 6
    (1996): 791, except that a significantly longer
    (1996): 791, except that a significantly longer
    reannealing hybridization was used."
BASE COUNT      0 172 a 162 c 209 g 127 t
ORIGIN
Query Match      12.4% Score 355.2; DB 13; Length 670;
Best Local Similarity 70.9%; Pred. No. 7e-57;
Matches 471; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 2195 AGAGAGGTGAAAGCGAGAACCTGGTGTCCGAGGTGCAATTGGATCAAAAGAGAAATCTG 2254
DB 6 AGAAGGTGAGAGGGAGAGCCAGGCATTCGAGGCATCAATGACAGAGGGAGATCCG 65
QY 2255 GGGTGGATGGCTTGATGGGGCCCGCAGGTCTTAAGGGGCAACCTGGGGATCCAGGTCTC 2314
DB 66 GTATCCAGGTCTGGTTGGACCTCTGGTGTACAGGTCTAGCCTGGAGACAGAGGGCCCC 125
QY 2315 AGGACCCCGCAGTTTGGATGGGAGCCCGAGAGAGTTTTCAGACAAATTTATTCGAC 2374
DB 226 CAGGACCACTGGATCGAATGGAACCTGCACGAGAAATTTTCAGAGAAATTTATTCGAC 185
QY 2375 AAGTTTGCACAGATGTAAATAGAGCCAGCTACCACTCTTACTTTCAGAGTGAAGAAATTA 2434
DB 186 AGGTGTGTTCAGATGTACTGAGAACCCAGTTACTCTGTGATCTCTCAGAGTGAAGGCTAC 245
QY 2435 GAATTTGTGATTCATTCCTGTCCCAACATGGTCTCCCGGGTATTTCTTGGGCCACCTGGTC 2494
DB 246 AAAATCTCAACCACTGTCTCAGTCCAGAGCGCTTCCCGAGGACTTCCAGGACCCAGGTC 305
QY 2495 CGATAGGCCAGAGGGTCCCGAGAGGATTACCTGTTTCCCGAGAGAGATGGTGTTCCTG 2554
DB 306 CAAGAGGCCAGAGAGGTCCCAAGAGGGTTCTCTGGTTTGGCCAGGGAATGACGGTGTTC 365
QY 2555 GATTAGTGGGTGTCTCTCGAGTCCAGGTGTTCAGAGGATTAAGAGGCTACCAAGAGAA 2614

```

```

Db      121  GAAACTGTCCAATTGGAAGTCCAAAAGCTACGAATCTACTGTGACCCAGAGCAGACAAC 180
QY      CCGGAGACAGCATGTGAGATCTCTGGATTT-----TGCCTTAATGTCCTCCAGTGAT 1299
Db      181  CCGGAGACAGCGTGTGAATACTCTGGATTTAATGCTGAGTGCCTCAACGGTCCCACTGAT 240
QY      1300  GTAGGTCAACTCCAGCTCCCTGCTGATTTGCTCCCGGAAAACCCAGGACTTCAAGGCCCC 1359
Db      241  GTAGGTCAACCGCAGCTCCCTGCTGATTTGCTCCAGGAAAACCCAGGACTTACAAAGGCC 300
QY      1360  AAGGTGACCTGACCTGCTCGGAAACCTGCTACCTCGGACAACTGCTCAAGATGCT 1419
Db      301  AAGGTGACCCCGACAGCTGCGAATCTCTGCTACCCCGGACGGCTGCTCAAGATGCT 360
QY      1420  AAGCTTGATATCAGGGAATTCAGGAGACACCAAGGTGTTCCAGGATCTCCAGGAATACAA 1479
Db      361  AAGCTTGATATCAGGGAATAGCAGGATCACCAAGGTGTTCCAGGATCTCCCGGAATACAA 420
QY      1480  GGAGCTCGAGGACTACAGGTTACAAAGGAGAACCCAGGCGAGATGTTGACAAAGGTGAT 1539
Db      421  GGAGCACAAAGGACTGCCAGGTTACAAAGGAGAACCTGGGAGAGATGGAGAAAAGGTTGAC 480
QY      1540  CGTGGACTTCTGCTTTCTCTGGCTTCATGGCATGCCAGGATCAAAAGGT 1590
Db      481  CGTGGACTTCTGCTTTCTCTGGCTTCATGGCATGCCAGGATCGAAGGT 531

RESULT 6
LOCUS   BE236040 518 bp mRNA linear EST 10-JUL-2000
DEFINITION 143702 MARC lPIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE236040
VERSION BE236040.1 GI:9020758
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 518)
AUTHORS Fahrenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,J., White,J.,
Vallet,J., Wise,T., Rohrer,G.A., Perte,G., Sultana,R., Quackenbush
J. and Keele,J.W.
TITLE Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
JOURNAL Mamm. Genome 13 (8), 475-478 (2002)
PUBMED 12226715
COMMENT CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 85 row: G column: 4
Seq primer: ATTTAGGTGACACTATAG.
FEATURES
source Location/Qualifiers
1..518
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC lPIG"
/note="Vector: pCMV SPORT6; Site.1: NotI; Site.2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 157 a 105 c 175 g 81 t

ORIGIN
Query Match 14.7%; Score 420.6; DB 10; Length 518;
Best Local Similarity 88.5%; Pred. No. 2.9e-69;
Matches 456; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1585 AAGGTGAAATGGTGCCAAAGGAGACAAAGGATCACCTGGATTTTATGCAAAAGGGT 1644
Db 4 AAGGTGAAATGGTCCCAAGAGGAGATAAAGGATCACCTGGATTTTATGCAAAAGGGT 63
QY 1645 GCARAAGGTGAAAAGGGGATGCTGGCTTCCTGGCTCCCTGCGATGCTGGGAGAACCA 1704
Db 64 GCGAAGGTGAAAAGGGGACTGTTGGTTCCTGGCTTCCTGGACGTGCTGGGAGAACCA 123
QY 1705 GGAAGACATGAAAAGGATGGATTAAATGGTAGTCCCGGTTTCAAGGGAGAGCAGGATCC 1764
Db 124 GGAAGACATGAAAAGGATGGATTAAATGGTAGTCCCGGTTTCAAGGGAGAGCAGGATCC 183
QY 1765 CCGTGTGCTCCGGGGCAGGATGGAACACGGGGAGAGCTTGGATCCAGGATTTTCCTGGA 1824
Db 184 CCAGGAGCCCGAGGGCAGGATGGATTGCGGGGAGAGCTTGGAAATCCAGGATTTTCCTGGA 243
QY 1825 AACCGAGGATTAAATGGGCCAAAAGGAGAAATTTGGGCTCCAGGACAGCAGGAGAAA 1884
Db 244 AACCGAGGATTAAATGGGACAAAAGGAGAAATTTGGGCTCCAGGACAGCAGGAGAAA 303
QY 1885 GGAGCCCGAGGATGCCCTGGTTTAAATGGGAAGCAATGGCTCACAGGCCAGCCTGGAACA 1944
Db 304 GGAGCTCCAGGGATACCTGGCTTGATGGAAGCGATGGGCCCGCCAGGTCCAGCCCGGAACA 363
QY 1945 CCGGATCTTAAGGAGAACCAAGGTGAACCTGGATTTCAAGGATGCTCCAGGATTTTCCTGGA 2004
Db 364 CCAGGACCTTAAGGAGAAATTAAGGTGAACCTGGACTCCAGGAGCTTCTGGG 423
QY 2005 CTCNAGGAGAACCCAGGAGAACCGGTTCCCGCAGGAGAACCCAGGATACATCGGTTTACCC 2064
Db 424 CTCNAGGAGAACCCAGGAGAGCTGGTCCCGCAGGAGAACCCAGGATACCTGGGCTTACCT 483
QY 2065 GCGATTCAGGAAAAAAGGGGACANAGGAATCA 2099
Db 484 GGAATTCAGGAAAAAAGGGGAGCAAAAGGAATCA 518

RESULT 7
LOCUS CA395709 527 bp mRNA linear EST 06-NOV-2002
DEFINITION cs68a05.Y1 Human Retinal pigment epithelium/choroid cDNA
(Un-normalized, unannotated); cs Homo sapiens cDNA clone cs68a05
5' mRNA sequence.
ACCESSION CA395709
VERSION CA395709.1 GI:24731429
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 527)
AUTHORS Wistow,G., Bernstein,S.L., Wyatt,M.K., Parris,R.N., Behal,A.,
Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
TITLE Expressed sequence tag analysis of human RPE/choroid for the
NEBank Project: Over 6000 non-redundant transcripts, novel genes
and splice variants
JOURNAL Mol. Vis. 8 (4), 205-220 (2002)
MEDLINE 22103460
PUBMED 12107410
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 68 row: a column: 35

```

```
RESULT 4
BX4311407      909 bp      mRNA      linear      EST 15-MAY-2003
LOCUS          BX4311407 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE009YK07
DEFINITION     S-PRIME, mRNA sequence.
ACCESSION      BX4311407
VERSION        BX4311407.1 GI:30773057
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 909)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91066 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3377.r for
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BAG0132B02_CS01162_1&cluster=3377.r.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/InvitrogenCorporation1600
Faraday Avenue Genoscope sequence ID : CS0BAG0132B02_CS01162_1.
FEATURES
Location/Qualifiers
1..909
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE009YK07"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT.6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT    251 a 201 c 205 g 249 t 3 others
ORIGIN
Query Match      15.2%; Score 436.8; DB 13; Length 909;
Best Local Similarity 99.3%; Pred. No. 2.7e-72;
Matches 438; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2425 GGAAGAAATTAGAAATTTGATCATTCCTCTCCACATGGCTCCCGGGTATTCCTGGG 2484
DB 30 GNAAGAAATTAGAAATTTGATCATTCCTCTCCACATGGCTCCCGGGTATTCCTGGG 89
QY 2485 CCACCTGGTCCGATAGCCCGACAGAGGTCCAGAGGATTACCTGGTTTGCAGGAAGAGAT 2544
DB 90 CCACCTGGTCCGATAGCCCGACAGAGGTCCAGAGGATTACCTGGTTTGCAGGAAGAGAT 149
QY 2545 GGTGTTCCCTGGATTAGTGGTGTCCCTGGACGTCCAGGTGTACAGGATTAAGGCCTA 2604
DB 150 GGTGTTCCCTGGATTAGTGGTGTCCCTGGACGTCCAGGTGTACAGGATTAAGGCCTA 209
QY 2605 CCAGGAAGAAATGGGAAAAGAGGAGCCAGAGGTTTGGGTATCCTCGAGAACAGGTCTCT 2664
DB 210 CCAGGAAGAAATGGGAAAAGAGGAGCCAGAGGTTTGGGTATCCTCGAGAACAGGTCTCT 269
QY 2665 CTTGGTCCCCCAGGTCCAGAGGCCCTCTCTGGAATTAAGCAAGAAGGTCTCTCAGAGAC 2724
DB 270 CTTGGTCCCCCAGGTCCAGAGGCCCTCTCTGGAATTAAGCAAGAAGGTCTCTCAGAGAC 329
QY 2725 CCAGGTCTCCCTGGCAAGATGAGACCATGGAACCTGGATCCAGGGGACACAGGC 2784
DB 330 CCAGGTCTCCCTGGCAAGATGAGACCATGGAACCTGGATCCAGGGGACACAGGC 389
QY 2785 CCCCCAGGCATCTGCCACCCATCACTATGTTTGTAGTGAATTCGCCAGAGAGATCGGTTTC 2844
```

```
Db 390 CCCCAGGCACTCGGACCCATCACTATGTTTGTAGTGTAATTCGCAAGAGATCCGTTTC 449
QY 2845 AGAAAGGACCAACTATTAG 2865
DB 450 AGAAAGGACCAACTATTAG 470

RESULT 5
BE899929
LOCUS          BE899929      531 bp      mRNA      linear      EST 25-APR-2001
DEFINITION     190456 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION      BE899929
VERSION        BE899929.1 GI:10387679
KEYWORDS       EST.
SOURCE         Bos taurus (cow)
ORGANISM       Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 531)
Smith, T.P.L., Grosse, M.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chikko-McKown, C.G.,
Perte, G., Holt, I., Karanycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
11282978
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 85 row: E column: 7
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1..531
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 20 and day 40
embryos."
BASE COUNT    151 a 120 c 146 g 104 t
ORIGIN
Query Match      14.7%; Score 421.4; DB 10; Length 531;
Best Local Similarity 88.7%; Pred. No. 2.1e-69;
Matches 471; Conservative 0; Mismatches 51; Indels 9; Gaps 1;
QY 1569 ACAGAACAGATGTGACTTTGTATATTGATGACCAACAAATTGAAACCAAGCCCTTACAT 1128
DB 1 ACAGAACAGATGTGACTTCTGTATATTGATGATCAACAAATTGAAACCAAGCCATTACAT 60
QY 1129 CCAGTTTAGGGATCTGTATCANTGGCAACCCCAATTCGAAATATTCTGGAAGAA 1188
DB 61 CCAGTTTAGGGATCTTTCATCAGTGGGCAACCCCAATTCGAAAGTACTCTCGGAAGAA 120
QY 1189 GAAACTGTTGAGTTGATGTCCTGAAAGTTCCGAAATCTACTGTGACCCAGAACAGAACAC 1248
```


Db 1920 GGTAGTCCGGTTTCAAGGAGAGCAGGATCCCTGGTGTCTCCGGGGCAGGATGGAACA 1979

Qy 1792 CGGGGAGAGCTGGATCCAGGATTTCTGGNAACCGAGGATTAATGGCCCAAGGGA 1851

Db 1980 CGGGGAGAGCTGGATCCAGGATTTCTGGNAACCGAGGATTAATGGCCCAAGGGA 2039

Qy 1852 GAAATTTGGGCTCCAGGACAGCAAGAAAGGAGCCCAAGGATGCTGGTTAAATG 1911

Db 2040 GAAATTTGGGCTCCAGGACAGCAAGAAAGGAGCCCAAGGATGCTGGTTAAATG 2099

Qy 1912 GGAAGCAATGGCTCAACAGCCAGCTTGAACACCGGATCTAAGGAAAGCAAGGTGAA 1971

Db 2100 GGAAGCAATGGCTCAACAGCCAGCTTGAACACCGGATCTAAGGAAAGCAAGGTGAA 2159

Qy 1972 CTTGGATTAAGAGGATGCTGGGCTTCTGGGCTCAAGGGAGACCAAGGACAGGCT 2031

Db 2160 CTTGGATTAAGAGGATGCTGGGCTTCTGGGCTCAAGGGAGACCAAGGACAGGCT 2219

Qy 2032 TCCCCAGGAGACCAAGGATACATGGGTTTACCGGGATTCAGGAAAGGAGGACAAA 2091

Db 2220 TCCCCAGGAGACCAAGGATACATGGGTTTACCGGGATTCAGGAAAGGAGGACAAA 2279

Qy 2092 GGAATCAAGGTGAAAGATTTTCAAGGCTCAAGGAGGAGGAGGAGGAGGAGGAT 2151

Db 2280 GGAATCAAGGTGAAAGATTTTCAAGGCTCAAGGAGGAGGAGGAGGAGGAGGAT 2339

Qy 2152 CAGGGCAACAGGAGATTTCAAGGCTCATGGTGCAGGAGGAGGAGGAGGAGGAGG 2211

Db 2340 CAGGGCAACAGGAGATTTCAAGGCTCATGGTGCAGGAGGAGGAGGAGGAGGAGG 2399

Qy 2212 GAACCTGGTGTCCAGGATGCTTGGATCAAGGAGGAGATCTGGGTTGGATGGCTGATG 2271

Db 2400 GAACCTGGTGTCCAGGATGCTTGGATCAAGGAGGAGATCTGGGTTGGATGGCTGATG 2459

Qy 2272 GGGCCGCGAGTCTTAAGGGGCAACCTGGGATCCAGGCTCCAGGAGCCCGCAGGTTG 2331

Db 2460 GGGCCGCGAGTCTTAAGGGGCAACCTGGGATCCAGGCTCCAGGAGCCCGCAGGTTG 2519

Qy 2332 GATGGGAAGCCCGAAGAGAGTTTTCAGAACAAATTTATTCGACAGTTTGACAGATGA 2391

Db 2520 GATGGGAAGCCCGAAGAGAGTTTTCAGAACAAATTTATTCGACAGTTTGACAGATGA 2579

Qy 2392 ATAAGAGCCAGCTACAGCTTACTTCAGAGTGGAGATTAAGAAATTTGATCATCTGC 2451

Db 2580 ATAAGAGCCAGCTACAGCTTACTTCAGAGTGGAGATTAAGAAATTTGATCATCTGC 2639

Qy 2452 CTGTCCCAACATGGCTCCCGGGTATTCCTGGGCAACCTGGTCCGATAGCCCGAGGCT 2511

Db 2640 CTGTCCCAACATGGCTCCCGGGTATTCCTGGGCAACCTGGTCCGATAGCCCGAGGCT 2699

Qy 2512 CCGAGAGATTAACCTGGTTTCCAGGAGAGATGGTGTCTCTGGATTAAGTGGTGTCTCT 2571

Db 2700 CCGAGAGATTAACCTGGTTTCCAGGAGAGATGGTGTCTCTGGATTAAGTGGTGTCTCT 2759

Qy 2572 GGAGCTCAGGTGTACAGGATTAAGAGGCTTACCAGGAAGAAATGGGGAAGAGGAGC 2631

Db 2760 GGAGCTCAGGTGTACAGGATTAAGAGGCTTACCAGGAAGAAATGGGGAAGAGGAGC 2819

Qy 2632 CAGGGTTTGGGTTTCTCGAGAACAAAGTCTCTGGTCCCGGATCCAGAGGAGGCTCT 2691

Db 2820 CAGGGTTTGGGTTTCTCGAGAACAAAGTCTCTGGTCCCGGATCCAGAGGAGGCTCT 2879

Qy 2692 CTTGGATTAAGCAAGAGTCTCCAGGAGAGCCAGGCTCTCCCTGGCAAGATGAGAC 2751

Db 2880 CTTGGATTAAGCAAGAGTCTCCAGGAGAGCCAGGCTCTCCCTGGCAAGATGAGAC 2939

Qy 2752 CATGGAAACCTCGAATCCAGGGAACCAAGGCCCCCAGGCACTCTCGCAACCCACTCA 2811

Db 2940 CATGGAAACCTCGAATCCAGGGAACCAAGGCCCCCAGGCACTCTCGCAACCCACTCA 2999

Qy 2812 TGTTTTAGTGTAAATGCGAAGAGATCCGTTTCAGAAAGAGCAAAACTATTAG 2865

Db 3000 TGTTTTAGTGTAAATGCGAAGAGATCCGTTTCAGAAAGAGCAAAACTATTAG 3053

RESULT 2

LOCUS BX458795 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE009YK07
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION BX458795
VERSION BX458795.1 GI:31033009
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Pelayes, D.
TITLE Full-length cDNA libraries and normalization.
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3377.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DE009AF04QPl&cluster=3377.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DE009AF04QPl.
Location/Qualifiers
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE009YK07"
/cissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 350 a 223 c 387 g 212 t 29 others

FEATURES

source
Query Match 28.8%; Score 825.6; DB 13; Length 1201;
Best Local Similarity 95.6%; Pred. No. 8.4e-146;
Matches 914; Conservative 1; Mismatches 30; Indels 11; Gaps 6;
Qy 1576 CCAGGATCAAGGGTGAATGGGTGCCAAGGAGACAAAGGATCCCTGGATTTATGGC 1635
Db 68 CCAGGATCAAGGGTGAATGGGTGCCAAGGAGACAAAGGATCCCTGGATTTATGGC 127
Qy 1636 AAAAAGGGTCAAAAGGTGAAAAGGGGAATGCTGGCTTCCCTGGCTCCCTGGACTGCT 1695
Db 128 AAAAAGGGTCAAAAGGTGAAAAGGGGAATGCTGGCTTCCCTGGCTCCCTGGACTGCT 187
Qy 1696 GGAGAACCAAGAGACATGGAAGAGATGGAATTAAATGGGTAGTCCCGTTTCAAGGAGAA 1755
Db 188 GGAGAACCAAGAGACATGGAAGAGATGGAATTAAATGGGTAGTCCCGTTTCAAGGAGAA 247
Qy 1756 GCAGGATCCCTGGTCTCCGGGAGGATGGAACACGGGAGAGAGCTCGAATCCAGGA 1815
Db 248 GCAGGATCCCTGGTCTCCGGGAGGATGGAACACGGGAGAGAGCTCGAATCCAGGA 307
Qy 1816 TTTCCTGGAACCCAGGATTAATGGGCCAAAAGGGAGAAATTTGGGCTCCAGGACAGAA 1875
Db 308 TTTCCTGGAACCCAGGATTAATGGGCCAAAAGGGAGAAATTTGGGCTCCAGGACAGAA 367
Qy 1876 GGAAAAAAGAGAGCCCAAGGATCCCTGGTTTAAATGGGAAGCAATGGCTCACAGGCCAG 1935
Db 368 GGAAAAAAGAGAGCCCAAGGATCCCTGGTTTAAATGGGAAGCAATGGCTCACAGGCCAG 427

Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 32 Row: k Column: 15
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18780272
 This clone has the following problem: retained intron.

Location/Qualifiers
 1. 4145
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4814199"
 /tissue_type="Brain, hippocampus"
 /clone_lib="NIH MGC_95"
 /lab_host="DH10B"
 /note="Vector: pBluescript"

BASE COUNT 1275 a 787 c 1017 g 1066 t

ORIGIN

Query Match 98.3%; Score 2815.2; DB 11; Length 4145;
 Best Local Similarity 99.3%; Pred. No. 0;
 Matches 2854; Conservative 0; Mismatches 3; Indels 17; Gaps 2;

QY 1 ATGGCTCAGTATATACATTTCTCTGCATGTTTGGTGGTCTCTCTCAGATTCGTG 60
 DB 188 ATGGCTCAGTATATACATTTCTCTGCATGTTTGGTGGTCTCTCTCAGATTCGTG 247

QY 61 TTAGCTCAAGATGGGGAAGTAAGATCAAGTTGTCGTAAGTCTCCGACAGATTAGTTTC 120
 DB 248 TTAGCTCAAGATGGGGAAGTAAGATCAAGTTGTCGTAAGTCTCCGACAGATTAGTTTC 307

QY 121 ATCTTAGATGGCTTTATAGTGTGGCCGAGAAAATTTGAATAGTGAAGTGGCTT 180
 DB 308 ATCTTAGATGGCTTTATAGTGTGGCCGAGAAAATTTGAATAGTGAAGTGGCTT 367

QY 181 GTCAATATCAAAAACCTTTGACATAGGCCCCAGTTTATTCAGTTGGAGTGTCAA 240
 DB 368 GTCAATATCAAAAACCTTTGACATAGGCCCCAGTTTATTCAGTTGGAGTGTCAA 427

QY 241 TATAGTGACTACCTGTGCTGGAGATTCCTCTCGAAGCTATGATTGAGGAGAACATTG 300
 DB 428 TATAGTGACTACCTGTGCTGGAGATTCCTCTCGAAGCTATGATTGAGGAGAACATTG 487

QY 301 ACGGACAGTGAATCCATCTACTTACGAGGAGAAACACAAAGACAGGAGGAGCCATC 360
 DB 488 ACGGACAGTGAATCCATCTACTTACGAGGAGAAACACAAAGACAGGAGGAGCCATC 547

QY 361 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCAGATTTCCTGACTAAGTAGCAGTG 420
 DB 548 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCAGATTTCCTGACTAAGTAGCAGTG 607

QY 421 GTACTTACGATGGCAATCCCAAGATGAGCTCAAGATGCAAGTCAAGCAGCAAGAT 480
 DB 608 GTACTTACGATGGCAATCCCAAGATGAGCTCAAGATGCAAGTCAAGCAGCAAGAT 667

QY 481 AGTAAGATACATTATTTGCTATTTGGTGGTTCAGAACAGAGATCCGACATTAGA 540
 DB 668 AGTAAGATACATTATTTGCTATTTGGTGGTTCAGAACAGAGATCCGACATTAGA 727

QY 541 GCTATTGCCAACAGCTTCCTCTACTTATGTTTATGTTGGAAGACTATATTGCAATA 600
 DB 728 GCTATTGCCAACAGCTTCCTCTACTTATGTTTATGTTGGAAGACTATATTGCAATA 787

QY 601 TCCAAATAAGGAGTGAAGACGAGAACTTTGTGAAGAACTCTGTCTCCACACGA 660
 DB 788 TCCAAATAAGGAGTGAAGACGAGAACTTTGTGAAGAACTCTGTCTCCACACGA 847

QY 661 ATCCAGTGGCAGCTCGTGATGAAGGGGATTGATATCTTTAGTATGTAAT 720

DB 848 ATTCAGTGGCAGCTCGTGATGAAGGGGATTTGATATTCTTTAGGTTAGATGTAAT 907
 QY 721 AAAAAGGTTAAGAAAAGATACAGCTTTACCAAAAAAGATATAAGGATATGAAAGTAA 780
 DB 908 AAAAAGGTTAAGAAAAGATACAGCTTTACCAAAAAAGATATAAGGATATGAAAGTAA 967
 QY 781 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTTCCCAAGAGTCTTCTCCATCA 840
 DB 968 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTTCCCAAGAGTCTTCTCCATCA 1027
 QY 841 TATGATTTGTTCTACTCAAAAGATTTAAAGTCAAGAAAATTTGGGATTTATGAGAATA 900
 DB 1028 TATGATTTGTTCTACTCAAAAGATTTAAAGTCAAGAAAATTTGGGATTTATGAGAATA 1087
 QY 901 TTAACATTTGATGGAAGGCCACAAATAGCAGTTACCTTAAATGGTGGGCAAAAACCTTA 960
 DB 1088 TTAACATTTGATGGAAGGCCACAAATAGCAGTTACCTTAAATGGTGGGCAAAAACCTTA 1147
 QY 961 TTATTTACAACAACCCAGCGTAATTAATGGCTCACAAGTGTACCTTTGCTAACCTTCAA 1020
 DB 1148 TTATTTACAACAACCCAGCGTAATTAATGGCTCACAAGTGTACCTTTGCTAACCTTCAA 1207
 QY 1021 GTTAAGAGCTTTGTTGATGAAGGCTGCACAAATTCGTCTCTTAGTAAACAGAACAGAT 1080
 DB 1208 GTTAAGAGCTTTGTTGATGAAGGCTGCACAAATTCGTCTCTTAGTAAACAGAACAGAT 1267
 QY 1081 GTGACTTTGTTATATTGATGACCAACAATTTGAAAACAAGCCCTTACATCCAGTTTAGGG 1140
 DB 1268 GTGACTTTGTTATATTGATGACCAACAATTTGAAAACAAGCCCTTACATCCAGTTTAGGG 1327
 QY 1141 ATCTTGATCAATGGCCAAACCCAAATTTGAAAATATTCTTGAAAAGAGAAACTGTTTCAG 1200
 DB 1328 ATCTTGATCAATGGCCAAACCCAAATTTGAAAATATTCTTGAAAAGAGAAACTGTTTCAG 1387
 QY 1201 TTTGATGTCCAAAAGTTGCGAATCTACTGTGACCCAGAACAGAACAAACCGGAGAGACGA 1260
 DB 1388 TTTGATGTCCAAAAGTTGCGAATCTACTGTGACCCAGAACAGAACAAACCGGAGAGACGA 1447
 QY 1261 TGTGAGATTCCTGGATTT-----TGCTTTAATGGTCCCACTGATGTAGGTTCAACT 1312
 DB 1448 TGTGAGATTCCTGGATTTAATGGAGAGTGGCTTAAATGGTCCCACTGATGTAGGTTCAACT 1507
 QY 1312 CCAGCTCCCTGTAATTGCTCTCCGGAAAACCCAGGACTTCAAGGCCCCCAAGGTGACCT 1371
 DB 1508 CCAGCTCCCTGTAATTGCTCTCCGGAAAACCCAGGACTTCAAGGCCCCCAAGGTGACCT 1567
 QY 1372 GGACTGCTGGAAACCTCTGGCTAGCTCCCTGGACCAACCTGGTCAAGATGGTAAAGCTGATAT 1431
 DB 1568 GGACTGCTGGAAACCTCTGGCTAGCTCCCTGGACCAACCTGGTCAAGATGGTAAAGCTGATAT 1619
 QY 1432 CAGGAAATTCAGGAGACACAGGTTTCCAGGATCTCCAGAAATACAGGAGCTCGAGGA 1491
 DB 1620 CAGGAAATTCAGGAGACACAGGTTTCCAGGATCTCCAGAAATACAGGAGCTCGAGGA 1679
 QY 1492 CTACAGGTTTACAAAGAGAAACCAAGGCGAGATGGTGACAAAGGATGATCTGGAATTCCT 1551
 DB 1680 CTACAGGTTTACAAAGAGAAACCAAGGCGAGATGGTGACAAAGGATGATCTGGAATTCCT 1739
 QY 1552 GGTCTTCTGGCTTCATGCTGCCAGGATCAAGGGTCAAAATGGTGGCCAAAGGAGAC 1611
 DB 1740 GGTCTTCTGGCTTCATGCTGCCAGGATCAAGGGTCAAAATGGTGGCCAAAGGAGAC 1799
 QY 1612 AAAGGATCACTGGATTTTATGCAAAAAGGGTGCAAAAGGTTGAAAGGGGAAATGCTGGC 1671
 DB 1800 AAAGGATCACTGGATTTTATGCAAAAAGGGTGCAAAAGGTTGAAAGGGGAAATGCTGGC 1859
 QY 1672 TTCCCTGCTCTCCCTGGACCTGTGAGAAACCGAGAGACATGGAAGGATGGAATTAATG 1731
 DB 1860 TTCCCTGCTCTCCCTGGACCTGTGAGAAACCGAGAGACATGGAAGGATGGAATTAATG 1919
 QY 1732 GGTAGTCCCGTTTCAAGGAGAGACAGGATCCCTGCTCGGGGAGGATGGAACA 1791

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 19:09:03 : Search time 5306 Seconds
(without alignments)
13123.318 Million cell updates/sec

Title: US-09-996-611b-5
Perfect score: 2865
Sequence: 1 aggttcactatattacatt.....gaaaggacaaactattag 2865

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562794

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_esthum:*
4: em_esthum:*
5: em_esthum:*
6: em_esthum:*
7: em_esthum:*
8: em_esthum:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2815.2	98.3	4145	11	BC045597
2	825.6	28.8	1201	13	BC458795
3	720	25.1	929	13	BQ893743
4	436.8	15.2	909	13	BC431407

5	421.4	14.7	531	10	BE899929
6	420.6	14.7	518	10	BE236040
7	377	13.2	527	14	CA395709
8	355.2	12.4	670	13	BU250509
9	344	12.0	744	10	BG699698
10	340.2	11.9	760	13	BU108124
11	309	10.8	327	14	N46880
12	291.2	10.2	2668	11	AK028339
13	291.2	10.2	2996	11	AK011547
14	291.2	10.2	3035	11	BC038017
15	289.5	10.1	3405	11	AK031071
16	280.5	9.8	2826	11	AK030350
17	277.3	9.7	387	14	N85186
18	276	9.6	276	9	AI434106
19	262.8	9.2	494	10	AW662485
20	255.2	8.9	357	14	N86630
21	251	8.8	4784	11	AK041115
22	250.8	8.8	4782	11	AK048546
23	250.4	8.7	452	9	AA329319
24	249.4	8.7	4783	11	AK029212
25	248.8	8.7	5157	11	AK040971
26	247.2	8.6	4018	11	AK031163
27	240.2	8.4	944	13	BU030399
28	240.2	8.4	5367	11	AK084803
29	238.4	8.3	481	9	AI127272
30	233	8.1	233	9	AA227783
31	231.8	8.1	464	14	T70851
32	220.8	7.7	3512	11	AK080682
33	206.6	7.2	3113	11	AK028295
34	201	7.0	2931	11	AK019448
35	200.8	7.0	2513	11	BC030649
36	198	6.9	4288	11	BC039222
37	195	6.8	1552	11	AK038481
38	189.4	6.6	3649	11	BC035387
39	186.8	6.5	3699	11	AK075707
40	185.6	6.5	1443	11	BC008620
41	184.6	6.4	485	9	AA451616
42	184.6	6.4	3216	11	AK053733
43	178.2	6.2	4287	11	AK044870
44	171.6	6.0	1144	14	CD496495
45	171.4	6.0	1190	11	AK012466

ALIGNMENTS

RESULT 1
BC045597
LOCUS BC045597 4145 bp mRNA linear HTC 04-MAR-2003
DEFINITION Homo sapiens, clone IMAGE:4814199, mRNA.
ACCESSION BC045597
VERSION BC045597.1 G1:28374134
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 4145)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Strausberg, R.
Direct Submission
Submitted (31-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 1A03, Bethesda, MD 20892-2590, USA

REMARK
CONTACT: MGC help desk
Email: cgabs@mail.nih.gov

Issue Procurement: Miklos Palkovits, M.D., Ph.D.
DNA Library Preparation: Michael J. Brownstein (NIGR1) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Qy 2813 GTTTTAGTGTAAATGCCAGAGAGATCCGTTCAGAAAGG 2852
 |||||
Dd 1375 GTTTTAAGTGTAATTGCCAGAGAGATCCGTTCAGAAAGG 1414

Search completed: October 2, 2003, 19:29:40
Job time : 724 secs

QY 541 GCTATTGCCAACAGCTTCGTCTACTATGTGTTTATATGTGGAGACTATATTGCAATA 600
 DB |||||||
 QY 736 GCTATTGCCAACAGCTTCGTCTACTATGTGTTTATATGTGGAGACTATATTGCAATA 795
 DB |||||||
 QY 601 TCCAAATTAAGGAAGTGAATGAAGCGAAACCTTTGTGAAGAATCTGTCTGTCCACACGA 660
 DB |||||||
 QY 796 TCCAAATTAAGGAAGTGAATGAAGCGAAACCTTTGTGAAGAATCTGTCTGTCCACACGA 855
 DB |||||||
 QY 661 ATTCCAGTGGCAGCTCGTGATGAAGGGGATTTGATATTTCTTTTAGTTTAGATGAAT 720
 DB |||||||
 QY 856 ATTCCAGTGGCAGCTCGTGATGAAGGGGATTTGATATTTCTTTTAGTTTAGATGAAT 914
 DB |||||||
 QY 721 AAAAAGTTTAAAGAAATACAGCTTTTCAACAAAAGATATAAGGATATGAAGTAACA 780
 DB |||||||
 QY 915 AAAAGTTTAAAGAAATACAGCTTTTCAACAAAAGATATAAGGATATGAAGTAACA 974
 DB |||||||
 QY 781 TCAAAAGTTTATCAGAACTCACAAGCAATGTTTTCAGAAAGTCTTCTCCATCA 840
 DB |||||||
 QY 975 TCAAAAGTTTATCAGAACTCACAAGCAATGTTTTCAGAAAGTCTTCTCCATCA 1034
 DB |||||||
 QY 841 TATGTAATTTGTCTACTCAAGATTTAAAGTCAAGAAAATTTGGGATTTATGGAAATA 900
 DB |||||||
 QY 1035 TATGTAATTTGTCTACTCAAGATTTAAAGTCAAGAAAATTTGGGATTTATGGAAATA 1094
 DB |||||||
 QY 901 TTAACATTATGAAGGCCA 921
 DB |||||||
 QY 1095 TTAACATTATGAAGGCCA 1115
 DB |||||||

RESULT 15

ABK71703

ID ABK71703 standard; cDNA; 1414 BP.

XX AC ABK71703;

XX XX

XX 30-JUL-2002 (first entry)

XX DE

XX Human dithp polynucleotide #169.

XX KW

Human; dithp; diagnostic and therapeutic polynucleotide; gene; ss; bone;
 cell proliferative disorder; cancer; tumour; autoimmune disorder; brain;
 inflammatory disorder; viral infection; bacterial infection; seizure;
 fungal infection; parasitic infections; developmental disorder; breast;
 endocrine disorder; metabolic disorder; neurological disorder; cervix;
 gastrointestinal disorder; transport disorder; gene therapy; kidney;
 adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;
 skin; testis; thymus.

XX KW

XX OS

XX XX

XX PN

XX W020020754-A2.

XX XX

XX 14-MAR-2002.

XX XX

XX 29-AUG-2001; 2001WO-US27127.

XX PR

XX 05-SEP-2000; 2000US-229747P.

XX PR

XX 05-SEP-2000; 2000US-229748P.

XX PR

XX 05-SEP-2000; 2000US-229749P.

XX PR

XX 05-SEP-2000; 2000US-229750P.

XX PR

XX 05-SEP-2000; 2000US-229751P.

XX PR

XX 05-SEP-2000; 2000US-230583P.

XX PR

XX 06-SEP-2000; 2000US-230505P.

XX PR

XX 06-SEP-2000; 2000US-230514P.

XX PR

XX 06-SEP-2000; 2000US-230515P.

XX PR

XX 06-SEP-2000; 2000US-230517P.

PR 06-SEP-2000; 2000US-230610P.
 PR 06-SEP-2000; 2000US-230865P.
 PR 06-SEP-2000; 2000US-230988P.
 PR 07-SEP-2000; 2000US-230951P.
 PR 07-SEP-2000; 2000US-231163P.
 PR 07-SEP-2000; 2000US-231167P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
 P1 Cones AL, Yu CY, Wright RJ, Gierzen D, Liu TP, Yap PE, Dahl CE;
 P1 Moniyam MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
 P1 Gerstin EH, Petalita CH, David MH, Panzer SR, Flores V, Daffo A;
 P1 Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
 XX
 XX WP: 2002-383054/41.
 DR P-PSDB; AB360112.

An isolated polynucleotide useful in diagnostics and therapeutics -

PS Claim 1; Page 456; 686pp; English.

XX The invention relates to human diagnostic and therapeutic (dithp)
 CC polynucleotides and their associated polypeptides (dithp polypeptides).
 CC The sequences of the invention are used in the treatment and diagnosis of
 CC cell proliferative disorders (e.g. atherosclerosis, cirrhosis), cancers
 CC (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,
 CC cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or
 CC thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,
 CC psoriasis, osteoporosis), viral infections, bacterial infections, fungal
 CC infections, parasitic infections, developmental disorders (e.g. anaemia,
 CC epilepsy), seizure disorders (e.g. thrombosis, aneurysm), metabolic disorders
 CC (e.g. obesity, diabetes), neurological disorders (e.g. stroke,
 CC amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal
 CC disorders (e.g. ulcerative colitis, lysinuria), and transport disorders
 CC (e.g. myotonic dystrophy, catatonias, peripheral neuropathy). Sequences
 CC ABK71535-ABK71809 represent human dithp polynucleotides of the invention.

XX Sequence 1414 BP; 425 A; 267 C; 295 G; 427 T; 0 other;

QY Query Match 15.8%; Score 453.6; DB 24; Length 1414;
 DB Best Local Similarity 99.1%; Pred No. 2.7e-107;
 DB Matches 456; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2393 TAAGAGCCGACGTACAGAGTCTTACTTCAGAGTGGAGAAATTGGAATTTGATCATGGC 2452
 DB |||||||
 DB 955 TGAAGCCCGAGTACCAGTCTTACTTCAGAGTGGAGAAATTAGAAATTTGATCATGGC 1014
 QY 2453 TGTCCTCAACATGGCTCCCGGGTATTCTTGGGCCACCTGGTCCGATAGAGCCAGAGGGTC 2512
 DB |||||||
 DB 1015 TGTCCTCAACATGGCTCCCGGGTATTCTTGGGCCACCTGGTCCGATAGAGCCAGAGGGTC 1074
 QY 2513 CCAGAGGATTACCTGGTTTCCAGGAGAGATGGTCTCTGATTAGTGGGTGTCCTCG 2572
 DB |||||||
 DB 1075 CCAGAGGATTACCTGGTTTCCAGGAGAGATGGTCTCTGATTAGTGGGTGTCCTCG 1134
 QY 2573 GACGTCCAGGTGTCAGAGGATTTAAAGGCTACCCAGGAGAAATGGGAAAAGAGAGCC 2632
 DB |||||||
 DB 1135 GACGTCCAGGTGTCAGAGGATTTAAAGGCTACCCAGGAGAAATGGGAAAAGAGAGCC 1194
 QY 2633 AAGGGTTTGGGTATCTCTGGAGAAACAAGTCTCTCTGCTCCCGAGGTCCAGAGGGCCCTC 2692
 DB |||||||
 DB 1195 AAGGGTTTGGGTATCTCTGGAGAAACAAGTCTCTCTGCTCCCGAGGTCCAGAGGGCCCTC 1254
 QY 2693 CTGGAATTAAGCAAGAGAGTCTCTCCAGGAGAGCCAGGTCTTCCCTGGCAAGATGGAGAC 2752
 DB |||||||
 DB 1255 CTGGAATTAAGCAAGAGAGTCTCTCCAGGAGAGCCAGGTCTTCCCTGGCAAGATGGAGAC 1314
 QY 2753 ATGGAAAACCTGGATCCCAAGGGCAACAGGCCCTCCAGGCATCTCCGACCATCACTAT 2812
 DB |||||||
 DB 1315 ATGGAAAACCTGGATCCCAAGGGCAACAGGCCCTCCAGGCATCTCCGACCATCACTAT 1374

OS	Homo sapiens.	CC	and as hybridisation probes for mapping naturally occurring genomic
XX	WO200240715-A2.	CC	sequences.
XX		XX	
XX	23-MAY-2002.	XX	
XX	06-SEP-2001; 2001WO-US27628.	XX	Sequence 1169 BP; 364 A; 224 C; 267 G; 314 T; 0 other;
XX	06-SEP-2000; 2000US-230505P.	XX	Query Match. 33.9%; Score 972.4; DS 24; Length 1169;
XX	06-SEP-2000; 2000US-230514P.	XX	Best Local Similarity 99.9%; Pred. No. 3e-242;
XX	06-SEP-2000; 2000US-230515P.	XX	Matches 973; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX	06-SEP-2000; 2000US-230517P.	XX	
XX	06-SEP-2000; 2000US-230518P.	XX	
XX	06-SEP-2000; 2000US-230519P.	XX	
XX	06-SEP-2000; 2000US-230595P.	XX	
XX	06-SEP-2000; 2000US-230597P.	XX	
XX	06-SEP-2000; 2000US-230598P.	XX	
XX	06-SEP-2000; 2000US-230599P.	XX	
XX	06-SEP-2000; 2000US-230610P.	XX	
XX	06-SEP-2000; 2000US-230655P.	XX	
XX	06-SEP-2000; 2000US-230988P.	XX	
XX	06-SEP-2000; 2000US-230989P.	XX	
XX	07-SEP-2000; 2000US-230951P.	XX	
XX	07-SEP-2000; 2000US-231163P.	XX	
XX	07-SEP-2000; 2000US-231167P.	XX	
XX	(INCY-) INCYTE GENOMICS INC.	XX	
XX	Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;	XX	
XX	Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;	XX	
XX	Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;	XX	
XX	Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;	XX	
XX	Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;	XX	
XX	WPI; 2002-527544/56.	XX	
XX	P-PSDB; ABP51293.	XX	
XX	Novel human disease detection and treatment polypeptide, useful in	XX	
XX	diagnosis, prevention or treatment of cell proliferative disorders e.g.	XX	
XX	arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder	XX	
XX	e.g. AIDS	XX	
XX	Claim 1; Page 341; 618pp; English.	XX	
XX	The invention relates to an isolated human disease detection and	XX	
XX	treatment (MDPT) polypeptide (I) selected from a polypeptide having a	XX	
XX	sequence selected from 254 sequences (ABP51231-ABP51484) given in the	XX	
XX	specification, a naturally occurring polypeptide comprising a sequence	XX	
XX	having at least 90% identity to (I) or a biologically active or	XX	
XX	immunogenic fragment of (I). (I) is useful for screening a compound for	XX	
XX	effectiveness as an agonist or antagonist, for screening a compound that	XX	
XX	specifically binds (I) or modulates the activity of (I), and for	XX	
XX	preparing a polyclonal or monoclonal antibody by hybridoma technology.	XX	
XX	Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for	XX	
XX	screening a compound for effectiveness in altering expression of a target	XX	
XX	polynucleotide comprising. Oligonucleotides and antibodies are useful for	XX	
XX	detecting MDPT in a sample or for assessing toxicity of a test compound,	XX	
XX	in a diagnostic test for a condition or a disease associated with the	XX	
XX	expression of MDPT in a biological sample, for detecting (I) in a sample,	XX	
XX	and for purifying (I) from a sample. A composition comprising (I), an	XX	
XX	agonist or antagonist is useful for treating a disease or condition	XX	
XX	associated with decreased or increased expression of functional MDPT.	XX	
XX	(I) or (II) are useful for diagnosing, treating or preventing disorders	XX	
XX	associated with aberrant expression of MDPT, where the disorders are	XX	
XX	selected from a cell proliferative disorder such as arteriosclerosis,	XX	
XX	cirrhosis, hepatitis, psoriasis, and cancer and an	XX	
XX	autoimmune/inflammatory disorder such as AIDS, Addison's disease,	XX	
XX	allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or	XX	
XX	rheumatoid arthritis. (II) are useful for creating knockin humanised	XX	
XX	animals or transgenic animals to model human diseases, in somatic or	XX	
XX	germline gene therapy, to generate a transcript image of a tissue or cell	XX	
XX	type, for detecting differences in the chromosomal location due to	XX	
XX	translocation or inversion among normal, carrier or affected individuals	XX	

PR 07-SEP-2000; 2000US-231163P.
 PR 07-SEP-2000; 2000US-231167P.
 PA (INCY-) INCYTE GENOMICS INC.
 XX Jackson S, Linceln SE, Altue CM, Dufour GE, Chalup MS, Hillman JL;
 PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
 PI Moniyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
 PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
 XX WPI; 2002-527544/56.
 DR P-PSDB; ABP51423.
 XX
 PT Novel human disease detection and treatment polypeptide, useful in
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
 PT e.g. AIDS
 XX
 PS Claim 1; Page 409-410; 618pp; English.
 XX
 CC The invention relates to an isolated human disease detection and;
 CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
 CC specification, a naturally occurring polypeptide comprising a sequence
 CC having at least 90% identity to (I) or a biologically active or
 CC immunogenic fragment of (I). (I) is useful for screening a compound for
 CC effectiveness as an agonist or antagonist, for screening a compound that
 CC specifically binds (I) or modulates the activity of (I), and for
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
 CC screening a compound for effectiveness in altering expression of a target
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
 CC detecting MDDT in a sample or for assessing toxicity of a test compound,
 CC in a diagnostic test for a condition or a disease associated with the
 CC expression of MDDT in a biological sample, for detecting (I) in a sample,
 CC and for purifying (I) from a sample. A composition comprising (I), an
 CC agonist or antagonist is useful for treating a disease or condition
 CC associated with decreased or increased expression of functional MDDT.
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders
 CC associated with aberrant expression of MDDT, where the disorders are
 CC selected from a cell proliferative disorder such as arteriosclerosis,
 CC cirrhosis, hepatitis, psoriasis, and cancer and an
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
 CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised
 CC animals or transgenic animals to model human diseases, in somatic or
 CC germline gene therapy, to generate a transcript image of a tissue or cell
 CC type, for detecting differences in the chromosomal location due to
 CC translocation or inversion among normal, carrier or affected individuals
 CC and as hybridisation probes for mapping naturally occurring genomic
 CC sequences.
 XX
 SQ Sequence 1184 BP; 368 A; 227 C; 272 G; 317 T; 0 other;
 Query Match 34.0%; Score 974; DB 24; Length 1184;
 Best Local Similarity 100.0%; Pred. No. 1.2e-242;
 Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGGCTCACTATTACATTTCTCTGCATGCTTTTGGTGGCTTCTTCAGAAATCTGTG 60
 DB 211 ATGGCTCACTATTACATTTCTCTGCATGCTTTTGGTGGCTTCTTCAGAAATCTGTG 270
 QY 61 TTAGCTGAAGTGGGAGTGAAGTCAAGTGTCTGCTACTGCTCGCAGAGATTAGTTTTC 120
 DB 271 TTAGCTGAAGTGGGAGTGAAGTCAAGTGTCTGCTACTGCTCGCAGAGATTAGTTTTC 330
 QY 121 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTGAATAGTGAAGTGGCTT 180
 DB 331 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTGAATAGTGAAGTGGCTT 390
 QY 181 GTCAATATCAAAAAACCTTTGACATAGGCGCGAGTTTATTCAGTTGGAGTGGTTCAA 240

DB 391 GTCAATATCAAAAAACCTTTGACATAGGCGCGAGTTTATTCAGTTGGAGTGGTTCAA 450
 QY 241 TATAGTCACTACCTGCTGCTGGAGATTCCTCTCGGAAGCTATGATTCAGGAGAAATTTG 300
 DB 451 TATAGTCACTACCTGCTGCTGGAGATTCCTCTCGGAAGCTATGATTCAGGAGAAATTTG 510
 QY 301 ACGGACAGCTGGAATCCCATCTACTCTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 360
 DB 511 ACGGACAGCTGGAATCCCATCTACTCTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 570
 QY 361 CAGTTTCGCTCGATTTACCTTTTTCGAAGTCTCTACGATTTCTGACTAAGATAGCAGTG 420
 DB 571 CAGTTTCGCTCGATTTACCTTTTTCGAAGTCTCTACGATTTCTGACTAAGATAGCAGTG 630
 QY 421 GTACTTACGATGSCAAATCCCAAGATGACGTCAAGATGCGAGCTCAACGACAGAGAT 480
 DB 631 GTACTTACGATGSCAAATCCCAAGATGACGTCAAGATGCGAGCTCAACGACAGAGAT 690
 QY 481 AGTAAGATAACATTTATTTGCTATTGCTTTGTTTTCAGAAACAGAACTCCGAACCTAGA 540
 DB 691 AGTAAGATAACATTTATTTGCTATTGCTTTGTTTTCAGAAACAGAACTCCGAACCTAGA 750
 QY 541 GCATTCCCAACAGCTTCGCTACTTATGCTTTTATGTTGGAAGACTATATTGCAATA 600
 DB 751 GCATTCCCAACAGCTTCGCTACTTATGCTTTTATGTTGGAAGACTATATTGCAATA 810
 QY 601 TCCAAATAGGGAAGTGAAGACAGAACTTTGTGAAGAACTCTCTGTGTCCACACCA 660
 DB 811 TCCAAATAGGGAAGTGAAGACAGAACTTTGTGAAGAACTCTCTGTGTCCACACCA 870
 QY 661 ATTCAGTGGCAGCTGTGATGAAAGGGGATTTGATATCTTTTAGTTTAGATGTAAT 720
 DB 871 ATTCAGTGGCAGCTGTGATGAAAGGGGATTTGATATCTTTTAGTTTAGATGTAAT 930
 QY 721 AAAAGGTTAAGAAAAGATACAGCTTCCACCAAAAGATAAAGATTAAGTAACA 780
 DB 931 AAAAGGTTAAGAAAAGATACAGCTTCCACCAAAAGATAAAGATTAAGTAACA 990
 QY 781 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTCCCAAGAGCTTCTCTCCATCA 840
 DB 991 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTCCCAAGAGCTTCTCTCCATCA 1050
 QY 841 TATGTTTGTGCTTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTAGGAGATA 900
 DB 1051 TATGTTTGTGCTTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTAGGAGATA 1110
 QY 901 TTAATCTTATGATGAAGGCCACAAATAGCAGTTTACCTTAAATGGTGTGCAAAAACTTTA 960
 DB 1111 TTAATCTTATGATGAAGGCCACAAATAGCAGTTTACCTTAAATGGTGTGCAAAAACTTTA 1170
 QY 961 TTATTTTCAACAAC 974
 DB 1171 TTATTTTCAACAAC 1184
 RESULT 13
 ABQ72511
 ID ABQ72511 standard; cDNA; 1169 BP.
 XX
 AC ABQ72511;
 XX
 DT 03-SEP-2002 (first entry)
 XX
 DE Human MDDT encoding cDNA SEQ ID NO 63.
 XX
 KW Human; MDDT; disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
 KW hepatotropic; antiinflammatory; antipsoriatic; cytotostatic; anti-HIV;
 KW antiallergic; antianaemic; antiasthmatic; antiatherosclerotic; antigout;
 KW neuroprotective; antirheumatic; antiarthritic; gene; ss.
 XX

PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -

PS Claim 1: SEQ ID NO 4177: 10078pp; English.

The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with neurotropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

CC C.N.S. disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

Sequence 2209 BP; 674 A; 413 C; 544 G; 577 T; 1 other;

Query Match 36.3%; Score 1039.8; DB 22; Length 2209;
Best Local Similarity 98.1%; Pred. No. 1.3e-259;
Matches 1107; Conservative 0; Mismatches 12; Indels 10; Gaps 5;

Qy

1747 AAGGGAGAAGCAGGATCCCTCGTGTCTCCGGGCAGGATGGAAACACGGGGAGAG-----C 1801
|||||
|||

Dd

4 AAGGGAGAAGCAGGATCCCTCGTGTCTCCGGGCAGGATGGAAAACACGGGGAGAGAACC 63
|||||
|||

Qy 1802 CTGGAATCCAGATT - - CCGTGGAAACCGAGGATTATGGG - CCAAAGGGAGAAATTG 1858
||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Dd 64 CTGGAATCCCCAGGATTCCCTGGAAACCGAGGATTATGGGCCCAAGGGAGAAATTG 123

Qy 1859 GGCCT - CCAGGACAGCAAGGAAAAAAGGAGCCCGGGATG - CCTGTTTAAATGGGAAG 1916
|||||
|||||
Db 124 GGCCTCCAGGACAGCAAGGAAAAAAGGAGCCCGGGATGCCCTGGTTTAATGGGAAG 183

QY 1917 CAATGGCTCACCAGGCCAGCCTGGAAACACCGGGATCTTAAGGGAGCAAGGTGAACCTGG 1976
|||||
Db 184 CAATGGCTCACCAGGCCAGCCTGGAAACACCGGGATCTTAAGGGAGCAAGGTGAACCTGG 243

QY	1977	AATTC AAGGG ATG CCT GGG GTT CTG GCT CAAG GGACA ACAG GAGCA ACG GTT CCCC	2036
Db	244	AATTCAAGGGATGCC TGGGGCTTCTGGGCTC AAGGGAGAAACAGGAGCAACGGGTCCCC	333

Qy	2037	AGGAGAACCAGGATACATGGTTTACCCTGGGATTCAAGGAAAAAAGGGGGACAAGAAGAA	2096
Db	304	AGGAGAACCAGGATACATGGTTTACCCTGGGATTCAAGGAAAAAAGGGGGACAAGAAGAA	363

Qy 2097 TCAAGGTGAAAAGGTTTTCAGGTCAAAAGGGAGAAAATGGHAGACAGGGAATTCACGG 2156

Db 364 TCAAGGTGAAAAGGTTTTCAGGTCAAAAGGGAGAAAATGGHAGACAGGGAATTCACGG 423

[illegible]

Qy 2217 TGGTGTCCGAGGTGCCATTGGATCAAAGGAGAACTCTGGGTGGATGGCTTGATGGGCCC 2276

Db 484 TGGTGTCCGAGGTGCCATTGGATCAAAGGAGAACTCTGGGTGGATGGCTTGATGGGCCC 543

Qy 2277 CGAGGTCCTAAGGGGCAACCTGGGGATCCAGGTCTCAGGGACCCCGAGGTTGGATGG 2336
|||
Db 544 CGAGGTCCTAAGGGGCAACCTGGGGATCCAGGTCTCAGGGACCCCGAGGTTGGATGG 603
|||

QY	2337	GAAGCCCGGAGAGAGATTTTCAGAACAAATTTATTCGACAGATTTGCACAGATGTAATAAG	2396
Db	604	GAAGCCCGGAGAGAGATTTTCAGAACAAATTTATTCGACAGATTTGCACAGATGTAATAAG	663

QY 2397 AGCCAGCTACCAAGTCTTACTTCAGAGTGGAGAAATTAGAATTGTGATCATTCGCTGTC 2456

664	Db	AGCCACGCTACCAAGTCTTACTTCAGAGTGGAGAAATTAGAAATTGTGATCATTGCCTGTC	723
2457	Qy	CCAACATGGCTCCCGGGTATTCTCTGGGCCACCTGTCGGATAGGCCACAGAGGTCCTCAG	2516
724	Db	CCAACATGGCTCCCGGGTATTCTGGGCCACCTGTCGGATAGGCCACAGAGGTCCTCAG	783
2517	Qy	AGGATTACCTGGTTTGGCCAGGAAGAGATGGTGTCTCTGGATTAGTGGGTCTCCCTGGACG	2576
784	Db	AGGATTACCTGGTTTGGCCAGGAAGAGATGGTGTCTCTGGATTAGTGGGTCTCCCTGGACG	843
2577	Qy	TCCAGGTGTCCAGAGGATTAAGAGGCTTACAGGAAGAAATGGGGAAGAAAGGAGGCAAGG	2636
844	Db	TCCAGGTGTCCAGAGGATTAAGAGGCTTACAGGAAGAAATGGGGAAGAAAGGAGGCAAGG	903
2637	Qy	GTTTGGGTATCTTGGGAACAAGGTCTCTCTGGTCTCCCGAGGTCCAGAGAGGCCCTCTGG	2696
904	Db	GTTTGGGTATCTTGGGAACAAGGTCTCTCTGGTCTCCCGAGGTCCAGAGAGGCCCTCTGG	963
2697	Qy	AATAAGCAAGAAGGTCTCCAGAGAGCCAGAGTCTCCCTGGCAAGATGGAGACATGG	2756
964	Db	AATAAGCAAGAAGGTCTCCAGAGAGCCAGAGTCTCCCTGGCAAGATGGAGACATGG	1023
2757	Qy	AAAACCTGGAAATCCAGGGCAACAGAGCCCCCAGGCAATCGGCACCCATCATGTGTT	2816
1024	Db	AAAACCTGGAAATCCAGGGCAACAGAGCCCCCAGGCAATCGGCACCCATCATGTGTT	1083
2817	Qy	TAGTGTAAATTGCCAAGAGATCGTTTCAGAAAGGACCAAACTATTAG	2865
1084	Db	TAGTGTAAATTGCCAAGAGATCGTTTCAGAAAGGACCAAACTATTAG	1132

RESULT 12
ABQ72640
ID ABQ7

ID ABQ72640 standard; cDNA; 1184 BP.

AC ABQ72640;

03-SEP-2002 (first entry)

XX
PE Human MONT encoding cDNA SEC ID NO 192

Human; MDDT; disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
 KW hepatotopic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;
 KW antiangiogenic; antianaemic; antilasthmatic; antiarteriosclerotic; antiglyc;
 KW neuroprotective; antirheumatic; antiarthritis; gene; ss.

YY
SO
OS
Homosapiens.

XX
WM300240715-A2

XX

XX
XX

PR 06-SEP-2000; 2000US-230514P.

PM 06-SEP-2000; 2000US-230317Z;
PM 06-SEP-2000; 2000US-230317Z;

PR 06-SEP-2000; 2000US-230518P.
PR 06-SEP-2000; 2000SI-230518P.
PR 06-SEP-2000; 2000SI-230518P.

PR 06-SEP-2000; 2000JS-230595d.

PR 06-SEP-2000; 2000US-230598P.

PR 06-SEP-2000; 2000US-230610P.

2R 06-SEP-2000; 2000US-230938P.

PR 07-SEP-2000; 2000US-23095-2
PR 07-SEP-2000; 2000US-23095-2

PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX Claim 1; Page 4838; 6221pp; English.
 XX The invention relates to polynucleotides (AAK51456-AAK51435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
 CC cytokine cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 SQ Sequence 2209 BP; 674 A; 413 C; 544 G; 577 T; 1 other;

Query Match 36.3%; Score 1039.8; DB 22; Length 2209;
 Best Local Similarity 98.1%; Pred. No. 1.3e-259;
 Matches 1107; Conservative 0; Mismatches 12; Indels 10; Gaps 5;

Qy 1747 AAGGGAGACGAGGATCCCTGGTGTCTCCGGGCGAGATGGAAACACGGGGAGAG-----C 1801
 Db 4 AAGGGAGACGAGGATCCCTGGTGTCTCCGGGCGAGGATGGAAACACGGGGAGAGAACCC 63
 Qy 1802 CTGGAATCCAGGATTT--CTCGAAACGAGGATTAATGGG--CMAAAGGGAGAAATG 1859
 Db 64 CTGGAATCCCGAGATTTCTCTGAAACCGAGGATTAATGGGCCCAAAAGGGAGAAATG 123
 Qy 1859 GGCT--CCAGGACAGCAAGGAAAGGAGGAGCCCGAGGATG--CTGGTTTAAATGGGAG 1916
 Db 124 GGCTCTCCAGGACAGCAAGGAAAGGAGGAGCCCGAGGATGCTGGTTTAAATGGGAG 183
 Qy 1917 CAATGGTCTACCGGCGAGCTCTGAAACACCGGATCTAAGGGAAGCAAAAGGTGAACCTGG 1976
 Db 184 CAATGGTCTACCGGCGAGCTCTGAAACACCGGATCTAAGGGAAGCAAAAGGTGAACCTGG 243
 Qy 1977 AATTCAGGATGCTGGGCTTCTGGCTCAGGAGGAGCAACGAGGAGACCGGTTCCCG 2036
 Db 244 AATTCAGGATGCTGGGCTTCTGGCTCAGGAGGAGCAACGAGGAGACCGGTTCCCG 303
 Qy 2037 AGGAGAACCGAGGATACATGGGTTTACCCGGGATTCAGGAAAGGAGGAGCAAAAGGAAA 2096
 Db 304 AGGAGAACCGAGGATACATGGGTTTACCCGGGATTCAGGAAAGGAGGAGCAAAAGGAAA 363
 Qy 2097 TCAAGGTGAAAAGGATTCAGGCTCAAAAGGAGGAGAAATGGAAGACAGGGAATCCAGG 2156
 Db 364 TCAAGGTGAAAAGGATTCAGGCTCAAAAGGAGGAGAAATGGAAGACAGGGAATTCAGG 423
 Qy 2157 GCAACAGGGAATTCAGGCTCATGGTGCAGGAGGAGAGAGGTTGAAAGGGAGAAC 2216
 Db 424 GCAACAGGGAATTCAGGCTCATGGTGCAGGAGGAGAGAGGTTGAAAGGGAGAAC 483
 Qy 2217 TGGTGTCCGAGTCCCATTTGATCAAAAGGAGAAATCTGGGTTGGATGGCTTATGGGGCC 2276
 Db 484 TGGTGTCCGAGTCCCATTTGATCAAAAGGAGAAATCTGGGTTGGATGGCTTATGGGGCC 543
 Qy 2277 CGCAGGTCTTAAGGGGCAACTGGGGATCCAGGTTCTCCAGGAGCCCGAGTTTGGATGG 2336
 Db 544 CGCAGGTCTTAAGGGGCAACTGGGGATCCAGGTTCTCCAGGAGCCCGAGTTTGGATGG 603
 Qy 2337 GAAGCCCGGAGAGAGTTTTCAGAACAAATTTATTCGACAAATTTGACACAGATGTAATAAG 2396
 Db 604 GAAGCCCGGAGAGAGTTTTCAGAACAAATTTATTCGACAAATTTGACACAGATGTAATAAG 663
 Qy 2397 AGCCAGGTACAGTCTTACTTTCAGAGTGAAGAAATTTAGAAATTTGATCATTCGCTGTC 2456
 Db 664 AGCCAGGTACAGTCTTACTTTCAGAGTGAAGAAATTTAGAAATTTGATCATTCGCTGTC 723

Qy 2457 CCAACATGGTCTCCCGGTAATTCCTGGGCCACCTGGTCCGATAGGCCCCAGAGGGTCCCGAG 2516
 Db 724 CCAACATGGTCTCCCGGTAATTCCTGGGCCACCTGGTCCGATAGGCCCCAGAGGGTCCCGAG 783
 Qy 2517 AGCATTTACCTGGTTTCCCGAGGAAGAGATGGTCTCTGGATTAGTGGGTGTCCTTGGACG 2576
 Db 784 AGCATTTACCTGGTTTCCCGAGGAAGAGATGGTCTCTGGATTAGTGGGTGTCCTTGGACG 843
 Qy 2577 TCAGGTGTCCAGAGGATTTAAAGGCTTACCAGGAAGAAATGGGGAAGGAGGAGCAAGG 2636
 Db 844 TCAGGTGTCCAGAGGATTTAAAGGCTTACCAGGAAGAAATGGGGAAGGAGGAGCAAGG 903
 Qy 2637 GTTTGGGTATCTCGGAGAAACAAGGTCTCTCTGTCCTCCCGAGTCCAGAGGGCCCTCTCTGG 2696
 Db 904 GTTTGGGTATCTCGGAGAAACAAGGTCTCTCTGTCCTCCCGAGTCCAGAGGGCCCTCTCTGG 963
 Qy 2697 AATAAGCAAGAGGTCCTTCCAGGAGACCCAGGTCTTCCCTGGCAAGATGGAAGCATATGG 2756
 Db 964 AATAAGCAAGAGGTCCTTCCAGGAGACCCAGGTCTTCCCTGGCAAGATGGAAGCATATGG 1023
 Qy 2757 ABAACCTGGATCCAGGAGGCAACCGGCCCCCGAGGCATCTCGACCCATCATCTATGTT 2816
 Db 1024 ABAACCTGGATCCAGGAGGCAACCGGCCCCCGAGGCATCTCGACCCATCATCTATGTT 1083
 Qy 2817 TAGTGTAAATTCGCCAGAGAGATCCGTTCAAGAAAGGAGCAAACTATTAG 2865
 Db 1084 TAGTGTAAATTCGCCAGAGAGATCCGTTCAAGAAAGGAGCAAACTATTAG 1132

RESULT 11

AA:60188
 ID AAT60188p,standard; cDNA; 2209 BP.
 XX AC AA:60188;
 XX DT 22-OCT-2001 (first entry)
 XX DE Human polynucleotide SEQ ID NO 4177.
 XX KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX OS Homo sapiens.
 XX PN WO200153312-A1.
 XX PD 26-JUL-2001.
 XX PF 26-DEC-2000; 2000WO-US34263.
 XX PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598442.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 XX PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac R;
 XX WP: 2001-442253/47.
 DR P-PSDB; XAN41032.
 XX

XX	WPI; 2001-442253/47.
DR	P-PSDB; AAM39246.
XX	
2T	Novel nucleic acids and polypeptides, useful for treating disorders
2T	such as central nervous system injuries -
XX	
PS	Claim 1; SEQ ID NO 605; 10078pp; English.
XX	
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemia and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
SQ	Sequence 2230 BP; 679 A; 409 C; 550 G; 592 T; 0 other;
	Query Match 37.1%; Score 1062; DB 22; Length 2230;
	Best Local Similarity 100.0%; Pred. No. 2.1e-265;
	Matches 1062; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1804 GGAATCCCGAGATTCTCTGGAAACCGAGGATAATTCGGGCCAAAAGGAGAAATTTGGGCCT 1863
Db	91 GGAATCCCGAGATTCTCTGGNAACCGAGGATTATTCGGCCAAAAGGAGAAATTTGGGCCT 150
QY	1864 CCAGCAGCACGAAGGAAAAAAGAGCCCCCAGGGATCCCTGGTTTTAATGGAGAACAATGGC 1923
Db	151 CCAGCAGCACGAAGGAAAAAAGAGCCCCCAGGGATCCCTGGTTTTAATGGAGAACAATGGC 210
QY	1924 TCACAGACGCCAGCTCGAACACCGGATCTTAAGGAGCAAGAGTGAACTCGAATTCAA 1983
Db	211 TCACAGACGCCAGCTCGAACACCGGATCTTAAGGAGCAAGAGTGAACTCGAATTCAA 270
QY	1984 GGGATGTCCTGGGGCTTCTGGGCTCAAGGGAGAACCCAGGACCAACGGGTTCGCCAGGAGAA 2043
Db	271 GGGATGTCCTGGGGCTTCTGGGCTCAAGGGAGAACCCAGGACCAACGGGTTCGCCAGGAGAA 330
QY	2044 CCAGGATACATGGGTTTACC CGGGATTCAAGGAAAAAAGGGGACAAAGGAATCAAGGT 2103
Db	331 CCAGGATACATGGGTTTACC CGGGATTCAAGGAAAAAAGGGGACAAAGGAATCAAGGT 350
QY	2104 GAAAAAGGTATTCAAGGGTCAAAGGGGAAAAATGGAGACAGGGAATTCAGGGCNACAG 2163
Db	391 GAARAAGGTATTCAAGGGTCAAAGGGGAAAAATGGAGACAGGGAATTCAGGGCNACAG 450
QY	2164 GGAATTCAGGCCATCATGGTGCAAAAGGAGAGAGAGGTGAAAGGGAGAACCTGGTGTC 2223
Db	451 GGAATTCAGGCCATCATGGTGCAAAAGGAGAGAGAGGTGAAAGGGAGAACCTGGTGTC 510
QY	2224 CCAGGTGCCATTGGATCAAAAGAGAAATCTGGGGTGGATGGCTTCATGGGGCCCGCAGGT 2293
Db	511 CGNGGTGCCATTGGATCAAAAGAGAAATCTGGGGTGGATGGCTTCATGGGGCCCGCAGGT 570
QY	2284 CCTAAGGGGCAACTCTGGGGATTCAGGTCTTCAGGGACCCCCCAGGTTTGGATGGGAAGCCC 2343
Db	571 CCTAAGGGGCAACTCTGGGGATTCAGGTCTTCAGGGACCCCCCAGGTTTGGATGGGAAGCCC 630
QY	2344 GGAAGAGAGTTTTTCAGACAAATTTATTTCGACAAAGTTTGCACAGATGTAAATAGAGCCCG 2403
Db	631 GGAAGAGAGTTTTTCAGACAAATTTATTTCGACAAAGTTTGCACAGATGTAAATAGAGCCCG 690
QY	2404 CTACCAAGCTCTTACTTTCAGAGTGGGAAGAAATTAGAAAATTTGTGATCATTTGCTGCCCAACAT 2463

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wehrman T, Goodrich R;
DR P-PSDB; AAM78912.
XX WPI; 2001-476283/51.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX Claim 1; Page 2132-2134; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX SQ Sequence 2230 BP; 679 A; 409 C; 550 G; 592 T; 0 other;

Query March 37.1%; Score 1062; DB 22; Length 2230;
Best Local Similarity 100.0%; Pred. No. 2.1e-265;
Matches 1062; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1804 GGAATCCAGGATTTCTCGAACCAGGATTAATGGGCCAAAGGGAGAAATTCGGCT 1863
DB 91 GGAATCCAGGATTTCTCGAACCAGGATTAATGGGCCAAAGGGAGAAATTCGGCT 150

QY 1864 CCAGGACAGCAAGGAAAAAAGAGCCCGCAGGATCCCTGGTTTAAATGGGAAGCAATGCG 1923
DB 151 CCAGGACAGCAAGGAAAAAAGAGCCCGCAGGATCCCTGGTTTAAATGGGAAGCAATGCG 210

QY 1924 TCACCGAGGCTGACACCGGATCTAAGGAGCAAGAGGTGACCTCGAATTCNA 1983
DB 211 TCACCGAGGCTGACACCGGATCTAAGGAGCAAGAGGTGACCTCGAATTCNA 270

QY 1984 GGGATGCTCGGGCTTCTGGCTCAAGGGAGAACACAGGACCAACGGTTTCCCGAGGAA 2043
DB 271 GGGATGCTCGGGCTTCTGGCTCAAGGGAGAACACAGGACCAACGGTTTCCCGAGGAA 330

QY 2044 CCAGGATACATGGTTTACCCGGATTCAGGAGAAAGGGGACAAAGGAATTCAGGT 2103
DB 331 CCAGGATACATGGTTTACCCGGATTCAGGAGAAAGGGGACAAAGGAATTCAGGT 390

QY 2104 GAAAGAGGATTCAGGCTCAAAAGGGAGAAATGGAAGACAGGAAATTCAGGGCAACAG 2163
DB 391 GAAAGAGGATTCAGGCTCAAAAGGGAGAAATGGAAGACAGGAAATTCAGGGCAACAG 450

QY 2164 GGAATTCAGGCTCATGTGCAAAAGGAGAGAGGTGAAAGGAGAACTGGTGTC 2223
DB 451 GGAATTCAGGCTCATGTGCAAAAGGAGAGAGGTGAAAGGAGAACTGGTGTC 510

QY 2224 CGAGTGCCATTTGGATCAAAAGGAGAAATCTGGGTGGATGGCTTGTATGGGGCCCGAGGT 2283
DB 511 CGAGTGCCATTTGGATCAAAAGGAGAAATCTGGGTGGATGGCTTGTATGGGGCCCGAGGT 570

QY 2284 CCTAAGGGGCAACCTCGGGATTCAGGTCCTCAGGACCCCGAGGTTTGGATGGAGGCC 2343
DB 571 CCTAAGGGGCAACCTCGGGATTCAGGTCCTCAGGACCCCGAGGTTTGGATGGAGGCC 630

QY 2344 GGAAGAGGTTTTTCAGAACCAATTTATTCGACAAAGTTTGCACAGATGTAATAGAGCCCG 2403
DB 631 GGAAGAGGTTTTTCAGAACCAATTTATTCGACAAAGTTTGCACAGATGTAATAGAGCCCG 690

QY 2404 CTACAGCTCTTACTTCAGAGTGGAGAAATTAGAAATCTGATCATTCCTCTCCACAT 2463
DB 691 CTACAGCTCTTACTTCAGAGTGGAGAAATTAGAAATCTGATCATTCCTCTCCACAT 750

QY 2464 GGCTCCCGGGTATTCTCGGGCCACCTGGTCCGATAGGCCCCAGAGGGTCCAGAGGATTA 2523
DB 751 GGCTCCCGGGTATTCTCGGGCCACCTGGTCCGATAGGCCCCAGAGGGTCCAGAGGATTA 810

QY 2524 CTTGGTTTCCAGGAGAGATGGTCTTCTGGATAGTGGGTGCTCCCTGGAGCTCCAGGT 2583
DB 811 CTTGGTTTCCAGGAGAGATGGTCTTCTGGATAGTGGGTGCTCCCTGGAGCTCCAGGT 870

QY 2584 GTCCAGAGGATTAAGAGGGCTTACACAGAAAGAAATGGGAAAGGGAGGCCAAGGGTTGGG 2643
DB 871 GTCCAGAGGATTAAGAGGGCTTACACAGAAAGAAATGGGAAAGGGAGGCCAAGGGTTGGG 930

QY 2644 TATCTCGAGAACAGGTCTCTCTGTCCTCCAGGTCCAGAGGGCCCTCTGGAAATAGC 2703
DB 931 TATCTCGAGAACAGGTCTCTCTGTCCTCCAGGTCCAGAGGGCCCTCTGGAAATAGC 990

QY 2704g AAAGAGGTCCTCCAGGAGACCCAGGTCTCTCTGTCCTGCAAGAGATGGAGACCATGGAACCT 2763
DB 991 AAAGAGGTCCTCCAGGAGACCCAGGTCTCTCTGTCCTGCAAGAGATGGAGACCATGGAACCT 1050

QY 2764 GGAATCCAGGGCAACAGGCCCGCCAGGCATCTCGACCCCATCACTATGTTTGTGTA 2823
DB 1051 GGAATCCAGGGCAACAGGCCCGCCAGGCATCTCGACCCCATCACTATGTTTGTGTA 1110

QY 2824 ATTGCCAGAGAGATCCGTTTCAGAAAGGACCAAACTATTAG 2865
DB 1111 ATTGCCAGAGAGATCCGTTTCAGAAAGGACCAAACTATTAG 1552

RESULT 9
AAI58402
ID AAI58402 standard; cDNA; 2230 BP.
XX AC AAI58402;
XX DT 22-OCT-2001 (first entry)
XX DE Human polynucleotide SEQ ID NO 605.
XX KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX OS Homo sapiens.
XX FN W0200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid
 CC arthritis and multiple sclerosis), muscular disorders, respiratory
 CC diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),
 CC reproductive disorders, gastrointestinal disorders, pulmonary disorders,
 CC cardiovascular disorders (e.g. congenital heart defects, Einstein's
 CC anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute
 CC kidney failure and end-stage renal disease), hyperproliferative disorders
 CC (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g.
 CC septic shock, bursitis and appendicitis), allergic reactions and
 CC conditions (e.g. asthma), blood related disorders (e.g. thrombosis,
 CC atherosclerosis and myocardial infarction) and cancerous diseases.
 CC Sequences ABX73173-ABX74167 represent human novel polynucleotides of the
 CC invention.
 XX
 SQ Sequence 2375 BP; 731 A; 435 C; 600 G; 609 T; 0 other;

Query Match 44.7%; Score 1282; DB 25; Length 2375;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1584 AAAGGTTGAATGGTGCACAAAGGAGCAAAAGGATCACTGGATTTTATGGCAAAAGGG 1643
 Db 9 AAAGGTTGAATGGTGCACAAAGGAGCAAAAGGATCACTGGATTTTATGGCAAAAGGG 68
 1644 TGCAAAAGGTCAAAGGGGATGCTGGCTTCCCTGGCTCCCTGGACCTGCTGGAAACC 1703
 Db 69 TGCAAAAGGTCAAAGGGGATGCTGGCTTCCCTGGCTCCCTGGACCTGCTGGAAACC 128
 1704 AGGAAGACATGAAAGGATGGAATTAATGGGTAGTCCCGTTTCAAGGGAGAACAGGATC 1763
 Db 129 AGGAAGACATGAAAGGATGGAATTAATGGGTAGTCCCGTTTCAAGGGAGAACAGGATC 188
 1764 CCCTGGTCTCCGGGGAGGATGGAACACGGGAGAGCTGGAATCCAGGATTTCCCTGG 1823
 Db 189 CCCTGGTCTCCGGGGAGGATGGAACACGGGAGAGCTGGAATCCAGGATTTCCCTGG 248
 1824 AAACCGAGGATTAATGGGCCAAAGGGAGAAATTTGGGCTCCAGGACAGCAAGGAAAAA 1893
 Db 249 AAACCGAGGATTAATGGGCCAAAGGGAGAAATTTGGGCTCCAGGACAGCAAGGAAAAA 308
 1884 AGGAGCCCCAGGATGCTGTTTAAATGGGAGCAATGGCTCACCAGGCGCCTGGAAC 1943
 Db 309 AGGAGCCCCAGGATGCTGTTTAAATGGGAGCAATGGCTCACCAGGCGCCTGGAAC 368
 1944 ACCGGGATCTAAGGGAAGCAAGGATGGAACCTGGAAATCAAGGGATGCTGGGCTTCTGG 2003
 Db 369 ACCGGGATCTAAGGGAAGCAAGGATGGAACCTGGAAATCAAGGGATGCTGGGCTTCTGG 428
 2004 GCTCAAGGGAGAACAGGAGCAACCGGTTTCCAGGAGAACAGGATACATGGTTTACC 2063
 Db 429 GCTCAAGGGAGAACAGGAGCAACCGGTTTCCAGGAGAACAGGATACATGGTTTACC 488
 2064 CGGGATCTCAAGGAAAAAGGGGACAAAGGAAATCAAGGTCAAAAGGATTTACAGGCTCA 2123
 Db 489 CGGGATCTCAAGGAAAAAGGGGACAAAGGAAATCAAGGTCAAAAGGATTTACAGGCTCA 548
 2124 AAAGGGAGAAATGGAAGACAGGAAATTCAGAGGCAACAGGAAATCAAGGCCATCATGG 2183
 Db 549 AAAGGGAGAAATGGAAGACAGGAAATTCAGAGGCAACAGGAAATCAAGGCCATCATGG 608
 2184 TGCAAAAGGAGAGAGGTGAAAGGGAGACCTGGTCCGAGGTGCCATTCGATCAAA 2243
 Db 609 TGCAAAAGGAGAGAGGTGAAAGGGAGACCTGGTCCGAGGTGCCATTCGATCAAA 668
 2244 AGGGAATCTGGGTGGATGGCTTATGGGGCCCGCAGGCTCTTAAGGGGCAACCTGGGGA 2303
 Db 669 AGGGAATCTGGGTGGATGGCTTATGGGGCCCGCAGGCTCTTAAGGGGCAACCTGGGGA 728
 2304 TCCAGGTCTCCAGGACCCCGCAGGTTTGGATGGGAGCCCGGAGAGAGTTTTCAGAAACA 2363

Db 729 TCAGAGTCTCCAGGGACCCCGAGGTTTGGATGGGAGCCCGGAGAGAGATTTCAGAAACA 788
 Qy 2364 ATTATTTCGACAAAGTTTGACACAGATGTAATAAGAGCCAGCTACCAAGTCTTACTTCAGAG 2423
 Db 789 ATTATTTCGACAAAGTTTGACACAGATGTAATAAGAGCCAGCTACCAAGTCTTACTTCAGAG 849
 Qy 2424 TGGAGAAATTAGAAATTTGATCAATTCGCTGTCACCAACATGGCTCCCGGGTATTCCTGG 2483
 Db 849 TGGAGAAATTAGAAATTTGATCAATTCGCTGTCACCAACATGGCTCCCGGGTATTCCTGG 908
 Qy 2484 GCCACCTTGGTCCGATAGGCGCCAGAGGGTCCAGAGGATTAACCTGGTTTCCAGGAGAGAGA 2543
 Db 909 GCCACCTTGGTCCGATAGGCGCCAGAGGGTCCAGAGGATTAACCTGGTTTCCAGGAGAGAGA 968
 Qy 2544 TGGTGTCTCTGGATTTAGTGGGTGTCCTCGACGTCCAGGTGTCAGAGGATTTAAAGGGCT 2603
 Db 969 TGGTGTCTCTGGATTTAGTGGGTGTCCTCGACGTCCAGGTGTCAGAGGATTTAAAGGGCT 1028
 Qy 2604 ACCAGGAAGAAATGGGGAAAAAGGAGCCAAAGGGTTTGGGTATCTCTGGAGAAACAAGGTCC 2663
 Db 1029 ACCAGGAAGAAATGGGGAAAAAGGAGCCAAAGGGTTTGGGTATCTCTGGAGAAACAAGGTCC 1088
 Qy 2664 TCCTGTGTCCTCCAGGTCCAGAGGCGCTCTCTCGAATAAGCAAAAGGATTCCTCCAGGAGA 2723
 Db 1089 TCCTGTGTCCTCCAGGTCCAGAGGCGCTCTCTCGAATAAGCAAAAGGATTCCTCCAGGAGA 1148
 Qy 2724 CCCAGGCTCTCCCTGGCAAGATGGAGACCATGGAACCTGGAATCCAAAGGCAACCCAGG 2783
 Db 1149 CCCAGGCTCTCCCTGGCAAGATGGAGACCATGGAACCTGGAATCCAAAGGCAACCCAGG 2808
 Qy 2784 CCCCCAGGATCTGGACCCCATCATATGTTTATGTAATTCGCCAGAGAGATCCGTT 2843
 Db 1209 CCCCCAGGATCTGGACCCCATCATATGTTTATGTAATTCGCCAGAGAGATCCGTT 1268
 Qy 2844 CAGAAAAGGACCAAACTATTAG 2865
 Db 1269 CAGAAAAGGACCAAACTATTAG 1290
 RESULT 8
 AAK52045
 ID AAK52045 standard; cDNA; 2230 BP.
 XX
 AC AAK52045;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 590.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2001:57190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04038.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.

Qy 2124 AAAGGGAAGAAATGGAGACAGGGAATTCAGGGCAACAGGGAATTCAGGCCATCATGG 2183
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 549 AAAGGGAAGAAATGGAGACAGGGAATTCAGGGCAACAGGGAATTCAGGCCATCATGG 608
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 2184 TGCAAAAGGAGAGAGAGGTGAAAGGAGAGAACCTGGTCCGAGGTGCCATTGGATCAAA 2243
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 609 TGCAAAAGGAGAGAGAGGTGAAAGGAGAGAACCTGGTCCGAGGTGCCATTGGATCAAA 668
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 2244 AGCAGATCTGGGGTGGATGGCTGTATGGGGCCGAGGTCTTAAGGGCAACCTGGGA 2303
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 669 AGGAGAACTGGGGTGGATGGCTGTATGGGGCCGAGGTCTTAAGGGCAACCTGGGA 728
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 2304 TCCAGGTCCTCAGGAGCCCCAGGTTGGATGGAGCCGAGGAGAGGTTTCAGAAAC 2363
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 729 TCCAGGTCCTCAGGAGCCCCAGGTTGGATGGAGCCGAGGAGAGGTTTCAGAAAC 788
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 2364 ATTATTCGACAAAGTTTGCACAGATGTAATAGAGCCAGCTACCAAGTCTTACTTCAGAG 2423
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 789 ATTATTCGACAAAGTTTGCACAGATGTAATAGAGCCAGCTACCAAGTCTTACTTCAGAG 848
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 2424 TGAAGAATTAGAAATGTGATCATTCCTGTCCCAACATGGCTCCCGGGTATTCCTGG 2483
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 849 TGGAGAAATAGAAAATGTGATCATTCCTGTCCCAACATGGCTCCCGGGTATTCCTGG 908
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 2484 GCCACCTGGTCCGATAGCCGAGAGGTCCGAGAGATTACCTGGTTGCCAGGAGAGA 2543
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 909 GCCACCTGGTCCGATAGCCGAGAGGTCCGAGAGATTACCTGGTTGCCAGGAGAGA 968
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 2544 TGGTGTCTCTGGATTTAGTGGGTGCTCCCTGACGTCCAGGTGTCAGAGGATTAAAGGCT 2603
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 969 TGGTGTCTCTGGATTTAGTGGGTGCTCCCTGACGTCCAGGTGTCAGAGGATTAAAGGCT 1028
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 2604 ACCAGGAAGAAATGGGGAAGGAGGAGGAGGTTGGGTATCTCGAGAGCAAGGTTCC 2663
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 1029 ACCAGGAAGAAATGGGGAAGGAGGAGGAGGTTGGGTATCTCGAGAGCAAGGTTCC 1088
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 2664 TCCTGTCTCCGAGTCCAGAGGCTCTCTGGAATAGCAAGAGGTTCTCCAGGAGA 2723
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 1089 TCCTGTCTCCGAGTCCAGAGGCTCTCTGGAATAGCAAGAGGTTCTCCAGGAGA 1148
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 2724 CCCAGGTCTCCTGGCAAGATGGAGACCATGAAACCTGGAATCAAGGGCAACGAG 2783
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 1149 CCCAGGTCTCCTGGCAAGATGGAGACCATGAAACCTGGAATCAAGGGCAACGAG 1208
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 2784 CCCCCAGGATCTGGACCATCATCTATGTTTATGTTAGTGAATTCGAGAGAGATCCGTT 2843
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 1209 CCCCCAGGATCTGGACCATCATCTATGTTTATGTTAGTGAATTCGAGAGAGATCCGTT 1269
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 2844 CAGAAAAGGACCAAACTATTAG 2865
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 1269 CAGAAAAGGACCAAACTATTAG 1290
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 7
 ABX73394
 ID ABX73394 standard; DNA; 2375 BP.

XX AC ABX73394;

XX DT 18-MAR-2003 (first entry)

XX DE Human novel polynucleotide #222.

XX KW Human; Gene; ds; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 KW haemostatic; antiarteriosclerotic.

OS Homo sapiens.

XX PN US2002132753-A1.
 XX PD 19-SEP-2002.
 XX XX 17-JAN-2001; 2001US-0764864.
 PR 31-JAN-2000; 2000US-179065P.
 PR 04-FEB-2000; 2000US-180628P.
 PR 28-JUN-2000; 2000US-214886P.
 PR 07-JUL-2000; 2000US-216647P.
 PR 07-JUL-2000; 2000US-216880P.
 PR 11-JUL-2000; 2000US-217487P.
 PR 11-JUL-2000; 2000US-217496P.
 PR 14-JUL-2000; 2000US-218290P.
 PR 26-JUL-2000; 2000US-220963P.
 PR 26-JUL-2000; 2000US-220964P.
 PR 14-AUG-2000; 2000US-224518P.
 PR 14-AUG-2000; 2000US-224519P.
 PR 14-AUG-2000; 2000US-225267P.
 PR 14-AUG-2000; 2000US-225268P.
 PR 14-AUG-2000; 2000US-225270P.
 PR 14-AUG-2000; 2000US-225447P.
 PR 14-AUG-2000; 2000US-225757P.
 PR 14-AUG-2000; 2000US-225758P.
 PR 22-AUG-2000; 2000US-226868P.
 PR 30-AUG-2000; 2000US-228924P.
 PR 01-SEP-2000; 2000US-229287P.
 PR 01-SEP-2000; 2000US-229343P.
 PR 01-SEP-2000; 2000US-229344P.
 PR 01-SEP-2000; 2000US-229345P.
 PR 05-SEP-2000; 2000US-229509P.
 PR 05-SEP-2000; 2000US-229513P.
 PR 06-SEP-2000; 2000US-231413P.
 PR 21-SEP-2000; 2000US-234223P.
 PR 21-SEP-2000; 2000US-234274P.
 PR 25-SEP-2000; 2000US-234977P.
 PR 27-SEP-2000; 2000US-235834P.
 PR 29-SEP-2000; 2000US-236327P.
 PR 29-SEP-2000; 2000US-236367P.
 PR 29-SEP-2000; 2000US-236368P.
 PR 29-SEP-2000; 2000US-236369P.
 PR 29-SEP-2000; 2000US-236370P.
 PR 02-OCT-2000; 2000US-236802P.
 PR 02-OCT-2000; 2000US-237037P.
 PR 02-OCT-2000; 2000US-237038P.
 PR 02-OCT-2000; 2000US-237039P.
 PR 02-OCT-2000; 2000US-237040P.
 PR 13-OCT-2000; 2000US-239935P.
 PR 20-OCT-2000; 2000US-240960P.
 PR 20-OCT-2000; 2000US-241785P.
 PR 20-OCT-2000; 2000US-241809P.
 PR 01-NOV-2000; 2000US-244617P.
 PR 17-NOV-2000; 2000US-249299P.
 PR 08-DEC-2000; 2000US-251856P.
 PR 08-DEC-2000; 2000US-251868P.
 PR 08-DEC-2000; 2000US-251869P.
 XX (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX PI Rosen CA, Ruben SM, Barash SC;
 XX WPI; 2003-147444/14.
 DR P-PSDB; ABUS5134.
 XX New polypeptides and nucleic acids, useful in gene therapy for
 PT treating, inhibiting or preventing e.g. neural immune system,
 PT muscular, respiratory, reproductive, gastrointestinal, pulmonary,
 PT cardiovascular or renal disorders -
 XX Claim 1; SEQ ID NO 232; 402bp; English.

```

PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX

```

```

DR WPI; 2001-488783/53.
DR P-PSDB; AAU16066.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 1; SEQ ID No 232; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention.

```

```

Query Match 44.7%; Score 1282; DB 22; Length 2375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1584 AAAGGTGAAATGGGTGCCAAAGGAGACAAAGGATCACTGGATTTTATGGCAAAAGGG 1643
DB 9 AAAGGTGAAATGGGTGCCAAAGGAGACAAAGGATCACTGGATTTTATGGCAAAAGGG 68
QY 1644 TGCAAAGGTGAAAGGGGAAATGCTGGCTTCCTCGCTCCCTGACCTGCTGGAGAAC 1703
DB 69 TGCAAAGGTGAAAGGGGAAATGCTGGCTTCCTCGCTCCCTGACCTGCTGGAGAAC 128
QY 1704 AGGAGACATGGAAGGATGGATTAATGGGTAGTCCCGTTTCAAGGGAGAGCAGGATC 1763
DB 129 AGGAGACATGGAAGGATGGATTAATGGGTAGTCCCGTTTCAAGGGAGAGCAGGATC 188
QY 1764 CCTGTGCTCGGGCAGGATGGAACACGGGGAGAGCTGGATCCAGGATTTCTTGG 1823
DB 189 CCTGTGCTCGGGCAGGATGGAACACGGGGAGAGCTGGATCCAGGATTTCTTGG 248
QY 1824 AAACCGAGGATTAATGGGCAAAAGGGAGAAATGGGCTCCAGGACAGAGGAAAAA 1883
DB 249 AAACCGAGGATTAATGGGCAAAAGGGAGAAATGGGCTCCAGGACAGAGGAAAAA 308
QY 1884 AGAGCCCCCAGGATGCTTGGTTTAATGGGAAGCAATGGCTCACAGCCAGCTTGGAAC 1943
DB 309 AGAGCCCCCAGGATGCTTGGTTTAATGGGAAGCAATGGCTCACAGCCAGCTTGGAAC 369
QY 1944 ACCGGATCTAAGGGAAGCAAGGTGAACCTGGATTCAGGGATGCTTGGGGTCTCTGG 2003
DB 369 ACCGGATCTAAGGGAAGCAAGGTGAACCTGGATTCAGGGATGCTTGGGGTCTCTGG 428
QY 2004 GCTCAAGGGAGAACCCAGGAGCAACGGGTTCCCGAGGAGAACAGGATACATGGTTTACC 2063
DB 429 GCTCAAGGGAGAACCCAGGAGCAACGGGTTCCCGAGGAGAACAGGATACATGGTTTACC 488
QY 2064 CGGGATTCAAGGAAAAAAGGGGGACAAAGGAAATCAAGGTGAAAAAGGTATTCAGGGTCA 2123
DB 489 CGGGATTCAAGGAAAAAAGGGGGACAAAGGAAATCAAGGTGAAAAAGGTATTCAGGGTCA 548

```

QY	901	TTAACTATTGATGGAAGGCCAACAAATAGCAGTTACCTTAAATGGTGTGGACAAATCTTTA	960	XX
DB	1015	TTAACTATTGATGG-ATGCCACAATAGCAGTTACCTTAAATGGTGTGGACAAATCTTTA	1073	PD
QY	961	TTATTACAAACACGAGCTTAATATGGCTCACAGTGGTTACCTTTGCTTAACCTCAA	1020	XX
DB	1074	TTATTACAAACACGAGCTTAATATGGCTCACAGTGGTTACCTTTGCTTAACCTCAA	1133	PF
QY	1021	GTTAGACGCTGTTTGTATGATGAGGCTGGACCAAAATTCGCTCTTAGTACAGAACAGAT	1080	XX
DB	1134	GTTAGACGCTGTTTGTATGATGAGGCTGGACCAAAATTCGCTCTTAGTACAGAACAGAT	1193	PR
QY	1081	GTGACTTTGTATATTGATGACCAACAAATTTGAAAAAAGCCCTTACATCCAGTTTATGGG	1140	PR
DB	1194	G*GACTTTGTATATTGATGACCAACAAATTTGAAAAAAGCCCTTACATCCAGTTTATGGG	1253	PR
QY	1141	ATCTTTGATCAATGGGCAACCCAAATTTGAAAAAATTTCTGAAAAAGAAACTGTTTCAAG	1200	PR
DB	1254	ATCTTTGATCAATGGGCAACCCAAATTTGAAAAAATTTCTGAAAAAGAAACTGTTTCAAG	1313	PR
QY	1201	TTTGATGTCACAAAGTTGGAATCTACTGTGACCCAGAACACAAACCGGAGACAGCA	1260	PR
DB	1314	TTTGATGTCACAAAGTTGGAATCTACTGTGACCCAGAACACAAACCGGAGACAGCA	1373	PR
QY	1261	TGTGAGATTCCTGGATTTTTCCTTAATGTGCTCCAGTGATGTAGGTTCAACTCCAGCTCCC	1320	PR
DB	1374	TGTGAGATTCCTGGATTTTTCCTTAATGTGCTCCAGTGATGTAGGTTCAACTCCAGCTCCC	1433	PR
QY	1321	TGTATTTGCTCCGGGAACCGGACCTTCAAGGCCCAAGGTCACCTGGACTGGCT	1380	PR
DB	1434	TGTATTTGCTCCGGGAACCGGACCTTCAAGGCCCAAGGTCACCTGGACTGGCT	1493	PR
QY	1381	GGGAACCTCGGTACCTCGGCAACCTGCTCAAGATGTTAAGCCT-----	1425	PR
DB	1494	GGGAACCTCGGTACCTCGGCAACCTGCTCAAGATGTTAAGCCT-----	1553	PR
QY	1426	-----GGATATCAGGGAATTCAGGGACACCAAGT	1455	PR
DB	1554	TTAGTTCATCTCCGGTATATCTCGGATTACAGGATATCAGGGAATTCAGGGACACCAAGT	1613	PR
QY	1456	GTTCAGGATCTCCAGGATACAGGACTCAGGACTACAGGTTACAAAGGAGAACCA	1515	PR
DB	1614	GTTCAGGATCTCCAGGATACAGGACTCAGGACTACAGGTTACAAAGGAGAACCA	1673	PR
QY	1516	GGCGGAGATGGTGACAAG	1533	PR
DB	1674	GGCGGAGATGGTGACAAG	1691	PR
RESULT 6				XX
AAS26053				PD
ID AAS26053 standard; cDNA; 2375 BP.				XX
AC AAS26053;				XX
XX 07-NOV-2001 (first entry)				XX
DE Human cDNA encoding a novel secreted protein. Seq ID 232.				XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;				XX
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;				XX
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;				XX
KW vulnerrary; secreted protein; rheumatoid arthritis;				XX
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;				XX
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;				XX
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;				XX
KW corneal infection; wound healing; epithelial cell proliferative;				XX
KW skin ageing; food additive; preservative; antiproliferative.				XX
OS Homo sapiens.				OS
XX WO200155122-A2.				XX
PN				PN

0

0

Db 1321 TGTATTGTCTCCGGGAAACACGAGCTTCAGGCCCCAAGGTGACCTGGACTGGCT 1380
QY 1381 GGAACCTGGCTACCTCGACCACTGGTCAAGATGGTAAGCCT----- 1425
Db 1381 GGAACCTGGCTACCTCGACCACTGGTCAAGATGGTAAGCCTGGTACTGAAAGC 1440
QY 1426 -----GGATATCAGGGAATTGCAGGACACAGGT 1455
Db 1441 TTAGTCATCTCCGGTATATCTGGATTACAGGATATCAGGGAATTCAGGGACACAGGT 1500
QY 1456 GTTCCAGGATCTCCAGGAATACAAAGAGCTTCGAGGACTACCAAGTTACAAAGAGAACCA 1515
Db 1501 GTTCCAGGATCTCCAGGAATACAAAGAGCTTCGAGGACTACCAAGTTACAAAGAGAACCA 1560
QY 1516 GGGGAGATGGTACAG 1533
Db 1561 GGGGAGATGGTACAG 1578
RESULT 5
AAK52265
ID AAK52265 standard; cDNA; 1691 BP.
XX AAK52265;
XX
DT 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 810.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX
XX Tang Y, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Xia Y;
XX Zhao QA, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX P-PSDS; AAM79132.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX
XX Claim 1; Page 2710-2711; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1691 BP; 532 A; 326 C; 386 G; 447 T; C other;
Query Match 51.1%; Score 1462.8; DB 22; Length 1691;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1530; Conservative 0; Mismatches 2; Indels 46; Gaps 2;
QY 1 ATGGCTCACTATATACATTTCTGTCATGGTTTGGTCTGCTTCTTCAGAAATCTCTG 60
Db 115 ATGGCTCACTATATATACATTTCTGTCATGGTTTGGTCTGCTTCTTCAGAAATCTCTG 174
QY 61 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGCTGCTACTGCTCCGACAGATTAGTTTC 120
Db 175 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGCTGCTACTGCTCCGACAGATTAGTTTC 234
QY 121 ATCTTAGATGGCTTTATAGTGTGCCCCAGAAAACTTTGAAATAGTGAAGTGGCTT 180
Db 235 ATCTTAGATGGCTTTATAGTGTGCCCCAGAAAACTTTGAAATAGTGAAGTGGCTT 294
QY 181 GTCAATATCAAAAAAATTTGACATAGGCGCCGAAAGTTTATTCAGATTGGAGTGGTCAA 240
Db 295 GTCAATATCAAAAAAATTTGACATAGGCGCCGAAAGTTTATTCAGATTGGAGTGGTCAA 354
QY 241 TATAGTACTACCTGCTGCTGGAGATTCCTTCGGAAGCTATGATTCAGGAGAACATTTC 300
Db 355 TATAGTACTACCTGCTGCTGGAGATTCCTTCGGAAGCTATGATTCAGGAGAACATTTC 414
QY 301 ACCGACAGCTGAATCCATCTCTACTAGGAGAAACACAAACAGAGGAGGCCATC 360
Db 415 ACCGACAGCTGAATCCATCTCTACTAGGAGAAACACAAACAGAGGAGGCCATC 474
QY 361 CAGTTTGGCTCGATTACCTTTTTCGCAAGTCTCTCAGCATTTCTGACTAAGATAGCAGT 420
Db 475 CAGTTTGGCTCGATTACCTTTTTCGCAAGTCTCTCAGCATTTCTGACTAAGATAGCAGT 534
QY 421 GTACTTACGGATGGCAATCCCAAGATGACGTCAGAGTGCAGCTCAAGCCACAGAGAT 480
Db 535 GTACTTACGGATGGCAAGTCCCAAGATGACGTCAGAGTGCAGCTCAAGCCACAGAGAT 594
QY 481 AGTAAGATAACATTATTTGCTATTTGCTGTTGCTTCAGAAACAGAAAGATGCCGAACTT 540
Db 595 AGTAAGATAACATTATTTGCTATTTGCTGTTGCTTCAGAAACAGAAAGATGCCGAACTT 654
QY 541 GCTATTGCCAAAGCCTTCTGCTACTTATGTTTATTTGTTGAGAGACTATATTGCAATA 600
Db 655 GCTATTGCCAAAGCCTTCTGCTACTTATGTTTATTTGTTGAGAGACTATATTGCAATA 714
QY 601 TCCAAATTAAGGAAGTGAATGAAGCAAACTTTTGAAGATCTGCTGTCACCAACAGA 660
Db 715 TCCAAATTAAGGAAGTGAATGAAGCAAACTTTTGAAGATCTGCTGTCACCAACAGA 774
QY 661 ATTCCAGTGGCAGCTCGTGAAGAGGGATTTGATTTCTTTTATAGTTAGATCTAAAT 720
Db 775 ATTCCAGTGGCAGCTCGTGAAGAGGGATTTGATTTCTTTTATAGTTAGATCTAAAT 834
QY 721 AAAAGGTTTAAAGAAAGATACAGCTTTTCCACAAAAGATTAAGAGATATGAAGTAA 780
Db 835 AAAAGGTTTAAAGAAAGATACAGCTTTTCCACAAAAGATTAAGAGATATGAAGTAA 894
QY 781 TCAAAAGTTGATTTATCAGAACTCAAGCAATGTTTCCCAAGAGGCTTCTCTCCATCA 840
Db 895 TCAAAAGTTGATTTATCAGAACTCAAGCAATGTTTCCCAAGAGGCTTCTCTCCATCA 954
QY 841 TATGTTATTTGTTGCTACTCAAGATTTAAAGTCAAGAAAAATTTGGGATTTATGGAGAATA 900
Db 955 TATGTTATTTGTTGCTACTCAAGATTTAAAGTCAAGAAAAATTTGGGATTTATGGAGAATA 1014

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.

XX Homo sapiens.

OS WO200157190-A2.

PN 09-AUG-2001.

PP 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-0693325.

PR 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI: 2001-476283/51.

DR P-PSDB; AAM80116.

XX Nucleic acids encoding polypeptides with cytokine-like activities,

PT useful in diagnosis and gene therapy -

XX Claim 1; Page 4980-4981; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666

CC (AAM80020) are omitted as the relevant pages from the sequence listing

CC were missing at the time of publication.

XX Sequence 1578 BP; 496 A; 294 C; 359 G; 429 T; 0 other;

Query Match 51.1%; Score 1464.4; DB 22; Length 1578;
 Best Local Similarity 97.0%; Pred. No. 0;
 Matches 1531; Conservative 0; Mismatches 1; Indels 46; Gaps 2;

QY 1 ATGGCTCACTATATACATTTCTCTGCATGTTTGGTGGTCTCTTCCAGAAATCTGTG 63

DB 2 ATGGCTCACTATATACATTTCTCTGCATGTTTGGTGGTCTCTTCCAGAAATCTGTG 61

QY 61 TTAGCTGAAGATGGGGAAGTATAGATCAAGTTGCTGCTCTCCGACAGATTTAGTTTC 120

DB 62 TTAGCTGAAGATGGGGAAGTATAGATCAAGTTGCTGCTCTCCGACAGATTTAGTTTC 121

QY 121 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTGAATAGTGAAGAGTGGCTT 180

DB 122 ACCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTGAATAGTGAAGAGTGGCTT 181

QY 181 GTCAATATCACAAAAACCTTGCATAGGCGGAGCTTTATTCAAGTTGGAGTGGTTCAA 240

DB 182 GTCAATATCACAAAAACCTTGCATAGGCGGAGCTTTATTCAAGTTGGAGTGGTTCAA 241

QY 241 TATAGTCACTACCTGTGCTGGAGATTCTCTCGAAGCTATGATTCAGGAGAACATTTG 300

DB 242 TATAGTCACTACCTGTGCTGGAGATTCTCTCGAAGCTATGATTCAGGAGAACATTTG 301

QY 301 ACGCAGCAGTGGAAATCCATACCTACTACTAGGAGGAAAACACAAAGACAGGGAAGGCCATC 360

DB 302 ACGCAGCAGTGGAAATCCATACCTACTACTAGGAGGAAAACACAAAGACAGGGAAGGCCATC 361

QY 361 CAGTTTGGCTCGATTACCTTTTGCACAGTCTCAGATTCTGACTAAGATAGCAGTG 420

DB 362 CAGTTTGGCTCGATTACCTTTTGCACAGTCTCAGATTCTGACTAAGATAGCAGTG 421

QY 421 GTACTTACGGATGCGAAATCCCAAGATGACGTCAAGGATGCGACTCAAGCAGCAGAGAT 480

DB 422 GTACTTACGGATGCGAAATCCCAAGATGACGTCAAGGATGCGACTCAAGCAGCAGAGAT 481

QY 481 AGTAAGATAACATATTTGCTATTTGGTGTGTTTCAAGAACAGAAAGATGCCGAACATAGA 540

DB 482 AGTAAGATAACATATTTGCTATTTGGTGTGTTTCAAGAACAGAAAGATGCCGAACATAGA 541

QY 541 GCTATTGCCCAACAAGCCTTCCTCTACTCTATGCTGTTTATGTGGAAGACTATATTGCAATA 600

DB 542 GCTATTGCCCAACAAGCCTTCCTCTACTCTATGCTGTTTATGTGGAAGACTATATTGCAATA 601

QY 601 TCCAAATAAGGGAAGTGAAGAGCAGAAACTTGTGAAGAATCTGTGTCCCAACACGA 660

DB 602 TCCAAATAAGGGAAGTGAAGAGCAGAAACTTGTGAAGAATCTGTGTCCCAACACGA 661

QY 661 ATTCCAGTGGCAGCTCGTGTGAAGGGGATTTGATATCTTTTGGTTAGATGTAAAT 720

DB 662 ATTCCAGTGGCAGCTCGTGTGAAGGGGATTTGATATCTTTTGGTTAGATGTAAAT 721

QY 721 AAAAAGTTAAGAAAGATACAGCTTTTCAACAAAAGATAAAAAGATATGAAGTAACA 780

DB 722 AAAAAGTTAAGAAAGATACAGCTTTTCAACAAAAGATAAAAAGATATGAAGTAACA 781

QY 781 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTCCCAAGAGTCTCTCCATCA 840

DB 782 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTCCCAAGAGTCTCTCCATCA 841

QY 841 TATCTATTGTGCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATCGAGAATA 900

DB 842 TATGTAATTGTGCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATCGAGAATA 901

QY 901 TTAATAATTGATGGAAGGCCACAAATAGCAGTTACCTTAAATGGTGTGCAAAAATCTTA 960

DB 902 TTAATAATTGATGGAAGGCCACAAATAGCAGTTACCTTAAATGGTGTGCAAAAATCTTA 960

QY 961 TTAATTTACAAACACCGCGTAATTAATGGCTCACAAGTGGTTACCTTTGCTAACCCCTCAA 1020

DB 961 TTAATTTACAAACACCGCGTAATTAATGGCTCACAAGTGGTTACCTTTGCTAACCCCTCAA 1020

QY 1021 GTTAAGACGTTGTTGATGAAGCTGGCACCAAAATTCGCTCTTAGTAACAGAACAGAT 1080

DB 1021 GTTAAGACGTTGTTGATGAAGCTGGCACCAAAATTCGCTCTTAGTAACAGAACAGAT 1080

QY 1081 GTGACTTTGTATATTGATGACCAAAATTTGAAAACAAGCCCTTACATCAAGTTTAGG 1140

DB 1081 GTGACTTTGTATATTGATGACCAAAATTTGAAAACAAGCCCTTACATCAAGTTTAGG 1140

QY 1141 ATCTTGATCAATGGGCAACCCCAAAATTTGAAAATAATTTCTGAAAAGAAAGAACTGTTAG 1200

DB 1141 ATCTTGATCAATGGGCAACCCCAAAATTTGAAAATAATTTCTGAAAAGAAAGAACTGTTAG 1200

QY 1201 TTTGATGTCCTCAAAAGTTGCGAATCTACTGTGACCCAGAACAGAACCAACCGGAGACAGA 1260

DB 1201 TTTGATGTCCTCAAAAGTTGCGAATCTACTGTGACCCAGAACAGAACCAACCGGAGACAGA 1260

QY 1261 TGTGAGATTCCTGATTTTSCCTTAAATGTTCCAGTGTATGAGTTCAACTCCAGCTCC 1320

DB 1261 TGTGAGATTCCTGATTTTSCCTTAAATGTTCCAGTGTATGAGTTCAACTCCAGCTCC 1320

QY 1321 TGTATTGTCTCGGGGAAAACCCAGGACTTTCAAGGCCCCCAAGAGTGAAGTGAAGTGAAGT 1380

XX Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 1; Page 4966; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAK78123-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAK80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 SQ Sequence 1578 BP; 496 A; 294 C; 359 G; 429 T; 0 other;

Query Match 51.1%; Score 1464.4; DB 22; Length 1578;
 Best Local Similarity 97.0%; Pred. No. 0;
 Matches 1531; Conservative 0; Mismatches 1; Indels 46; Gaps 2;

QY 1 ATGGCTCACTATATTACATTCTCGCATGGTTTTCGCTGCTCTTCAGAAATCTGTG 60
 DB 2 ATGGCTCACTATATTACATTCTCGCATGGTTTTCGCTGCTCTTCAGAAATCTGTG 61
 QY 61 TTAGCTGAAGATCGGGAAGTAAGATCAAGTTGCTGCTCTCCGACAGATTAGTTTTC 120
 DB 62 TTAGCTGAAGATCGGGAAGTAAGATCAAGTTGCTGCTCTCCGACAGATTAGTTTTC 121
 QY 121 ATCTTAGAGTGGCTCTTATAGTGTGGCCAGAAACTTTGAATAGTGAAGTGGCTT 180
 DB 122 ATCTTAGAGTGGCTCTTATAGTGTGGCCAGAAACTTTGAATAGTGAAGTGGCTT 181
 QY 181 GTCAATATCAAAAAAATTTGACATAGGCGCGAAGTTTATCAAGTTGGAGTGTCAA 240
 DB 182 GTCAATATCAAAAAAATTTGACATAGGCGCGAAGTTTATCAAGTTGGAGTGTCAA 241
 QY 241 TATAGTACATACCTGTGTGGAGATTCCTCTCGGAAGTATGATTCAGAGAAATTTG 300
 DB 242 TATAGTACATACCTGTGTGGAGATTCCTCTCGGAAGTATGATTCAGAGAAATTTG 301
 QY 301 ACGGCAGCTGGAATCCATCTCTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 360
 DB 302 ACGGCAGCTGGAATCCATCTCTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 361
 QY 361 CAGTTTGGCTCGCATACCTTTTGGCAAGTCTCAGATTCTGTACTAAGATAGCAGTG 420
 DB 362 CAGTTTGGCTCGCATACCTTTTGGCAAGTCTCAGATTCTGTACTAAGATAGCAGTG 421
 QY 421 GTACTTACGATGCGAATCCCAAGATGAGCTCAAGATGCGCTCAAGCAGCAGAGAT 480
 DB 422 GTACTTACGATGCGAATCCCAAGATGAGCTCAAGATGCGCTCAAGCAGCAGAGAT 481
 QY 481 AGTAAGATAACATTATTTGCTATTTGGTTGGTTTCAGAAACAGAAAGATCCGAATAGA 540
 DB 482 AGTAAGATAACATTATTTGCTATTTGGTTGGTTTCAGAAACAGAAAGATCCGAATAGA 541
 QY 541 GCTATTGCCAACAGCCCTCGCTACTTATGTGTTTATGTGGAAGACTATATTCGAATA 600
 DB 542 GCTATTGCCAACAGCCCTCGCTACTTATGTGTTTATGTGGAAGACTATATTCGAATA 601
 QY 601 TCCAAAATAAGGGAAGTATGAAGACAGAACTTTGTGAAGATCTCTGTCTGCTCAACAGA 660
 DB 602 TCCAAAATAAGGGAAGTATGAAGACAGAACTTTGTGAAGATCTCTGTCTGCTCAACAGA 661
 QY 661 ATTCCAGTGGCAGCTCGTGTATGAAGGGGATTTGATTTCTTTTAGGTTTAGATGTAAT 720

DB 662 ATTCAGTGGCAGCTCGTGTATGAAGGGATTTGATATCTTTAGGTTTAGATGTAAT 721
 QY 721 AAAAAAGTTAAGAAAAGATAAGCTTTTCCACAAAAGATAAAAAGATATGAAGATAACA 780
 DB 722 AAAAAAGTTAAGAAAAGATAAGCTTTTCCACAAAAGATAAAAAGATATGAAGATAACA 781
 QY 781 TCAAAAAGTTGATTTATCAGAACTCTCAAGCAATGTTTCCCAAGAGGCTTCTCTCCATCA 840
 DB 782 TCAAAAAGTTGATTTATCAGAACTCTCAAGCAATGTTTCCCAAGAGGCTTCTCTCCATCA 841
 QY 841 TATGTATTTGTCTCTACTCAAAAGATTTAAAGTCAAGAAAATTTGGGATTTATGGAGAATA 900
 DB 842 TATGTATTTGTCTCTACTCAAAAGATTTAAAGTCAAGAAAATTTGGGATTTATGGAGAATA 901
 QY 901 TTAACCTATTGATGGAGGCCACAAATAGCAGTTACCTTAAATGTTGTGCGAAGAAATCTTA 960
 DB 902 TTAACCTATTGATGG-ATGCCCAAAATAGCAGTTACCTTAAATGTTGTGCGAAGAAATCTTA 960
 QY 961 TTAATTTACAAACACAGCGTAAATTAATGCTCACAAGTGGTTTACCTTTGCTTAACCTCAA 1020
 DB 961 TTAATTTACAAACACAGCGTAAATTAATGCTCACAAGTGGTTTACCTTTGCTTAACCTCAA 1020
 QY 1021 GTTAAGACGTTGTTTGTATGAGGCTGGCACCACAAATTCGCTCTTATGTAACAGAAACAGAT 1080
 DB 1021 GTTAAGACGTTGTTTGTATGAGGCTGGCACCACAAATTCGCTCTTATGTAACAGAAACAGAT 1080
 QY 1081 GTGACTTTGTATATTGATGACCAACAAATTTGAAAACAAAGCCCTTACATCCAGTTTATGGG 1140
 DB 1081 GTGACTTTGTATATTGATGACCAACAAATTTGAAAACAAAGCCCTTACATCCAGTTTATGGG 1140
 QY 1141 ATCTTGATCAATGGGCAAAACCCAAATTTGAAAATATCTTGGAAAAGAAAGAACTTTTCAG 1200
 DB 1141 ATCTTGATCAATGGGCAAAACCCAAATTTGAAAATATCTTGGAAAAGAAAGAACTTTTCAG 1200
 QY 1201 TTTGATGTCNAAAGTTCGGAATCTACTGTGACCCAGACAGAAACAAACCGGAGACAGCA 1260
 DB 1201 TTTGATGTCNAAAGTTCGGAATCTACTGTGACCCAGACAGAAACAAACCGGAGACAGCA 1260
 QY 1261 TGTGAGATCTCTGGATTTTGCCTTAAATGCTCCAGTGTAGTTCAGTTTCAACTCCAGCTCCC 1320
 DB 1261 TGTGAGATCTCTGGATTTTGCCTTAAATGCTCCAGTGTAGTTCAGTTTCAACTCCAGCTCCC 1320
 QY 1321 TGTATTTGTCTCCGGGAAAACACAGACTTCAAGGCCCCAAAGGTTGACCTGGAGTGGCT 1380
 DB 1321 TGTATTTGTCTCCGGGAAAACACAGACTTCAAGGCCCCAAAGGTTGACCTGGAGTGGCT 1380
 QY 1381 GGGAAACCTTGGCTACCTCGCAACCTGCTCAAGATGTTGTTAGGTTGTTGAGTACTGAAAGC 1425
 DB 1381 GGGAAACCTTGGCTACCTCGCAACCTGCTCAAGATGTTGTTAGGTTGTTGAGTACTGAAAGC 1440
 QY 1426 -----GGATATCAGGGAATTCAGGGGACACACAGGT 1455
 DB 1441 TTAGTCATCTCCGTTATATCTGGGATTACAGGATATCAGGGAACTGCAAGGACACAGGT 1500
 QY 1456 GTTCCAGGATCTCCAGGAATACAGGAGCTCCAGAGCTACAGGTTTACAAAGGAGAACCA 1515
 DB 1501 GTTCCAGGATCTCCAGGAATACAGGAGCTCCAGAGCTACAGGTTTACAAAGGAGAACCA 1560
 QY 1516 GGGCGAGATGGTGACAAG 1533
 DB 1561 GGGCGAGATGGTGACAAG 1578

RESULT 4
 AAK53249
 ID AAK53249 standard; cDNA; 1578 BP.
 XX
 AC AAK53249;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 2778.
 XX

181 GTCAATATCAAAAAAATTTGACATAGGGCCGAGCTTTATTTCAAGTTGGAGTGGTTCAA 240
 Db |||||
 182 GTCAATATCAAAAAAATTTGACATAGGGCCGAGCTTTATTTCAAGTTGGAGTGGTTCAA 241
 Qy |||||
 241 TATAGTGAATACCTGTGTGGAGATTCCTCTCGGAAGCTATGATTCAGGAGAAATTTG 300
 Db |||||
 242 TATAGTGAATACCTGTGTGGAGATTCCTCTCGGAAGCTATGATTCAGGAGAAATTTG 301
 Qy |||||
 301 ACGGACAGTGAATTCATCTACTCTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 360
 Db |||||
 302 ACGGACAGTGAATTCATCTACTCTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 361
 Qy |||||
 361 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCAGGATTTCTGACTAAGATAGCAGTG 420
 Db |||||
 362 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCAGGATTTCTGACTAAGATAGCAGTG 421
 Qy |||||
 421 GTACTTACGGATGGCAAAATCCCAAGATGACGTCAAGGATGCGAGTCAAGCAGCAAGAGAT 480
 Db |||||
 422 GTACTTACGGATGGCAAAATCCCAAGATGACGTCAAGGATGCGAGTCAAGCAGCAAGAGAT 481
 Qy |||||
 481 AGTAAGATACATTTATTCGTTATTTGGTGTGGTTCAGAAACAGAGATGCCGAATAGA 540
 Db |||||
 482 AGTAAGATACATTTATTCGTTATTTGGTGTGGTTCAGAAACAGAGATGCCGAATAGA 541
 Qy |||||
 541 GCTATTGCCAACAGCCTTCGTCTACTTATGTGTGTATSTGGAAGACTATATTGCAATA 600
 Db |||||
 542 GCTATTGCCAACAGCCTTCGTCTACTTATGTGTGTATSTGGAAGACTATATTGCAATA 601
 Qy |||||
 601 TCCAAATAGGGAAGTGAAGAGCAAACTTTGTGAAGATCTGTCTGTCACACAGA 660
 Db |||||
 602 TCCAAATAGGGAAGTGAAGAGCAAACTTTGTGAAGATCTGTCTGTCACACAGA 661
 Qy |||||
 661 ATTCCAGTGCACCTCGTGATGAAGGGATTTGATTTCTTTTAGTTTACATGTAAAT 720
 Db |||||
 662 ATTCCAGTGCACCTCGTGATGAAGGGATTTGATTTCTTTTAGTTTACATGTAAAT 721
 Qy |||||
 721 AAAAAAGTTAAGAAAAAGAAATACAGCTTTCACCAAAAAGATAAAAGGATATGAAGTAACA 780
 Db |||||
 722 AAAAAAGTTAAGAAAAAGAAATACAGCTTTCACCAAAAAGATAAAAGGATATGAAGTAACA 781
 Qy |||||
 781 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTTCCAGAGGTTCTCTCATCA 840
 Db |||||
 782 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTTCCAGAGGTTCTCTCATCA 841
 Qy |||||
 841 TATGTATTTGTGTCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGAATA 900
 Db |||||
 842 TATGTATTTGTGTCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGAATA 901
 Qy |||||
 901 TTAACATTTGATGGAAAGGCCACAATAGCAGTTACCTTAAATGGTGTGACAAAATCTTA 960
 Db |||||
 902 TTAACATTTGATGGAAAGGCCACAATAGCAGTTACCTTAAATGGTGTGACAAAATCTTA 961
 Qy |||||
 961 TTATTTTACAACACACAGCGTAATTAATGGCTCAAGTGTGTACCTTTGCTAACCTCAA 1020
 Db |||||
 962 TTATTTTACAACACACAGCGTAATTAATGGCTCAAGTGTGTACCTTTGCTAACCTCAA 1021
 Qy |||||
 1021 GTTAAGACGTTGTTGATGAAGCTCGCCACCAAAATTCGTCTCTTAGTAACAGAAACAGAT 1080
 Db |||||
 1022 GTTAAGACGTTGTTGATGAAGCTCGCCACCAAAATTCGTCTCTTAGTAACAGAAACAGAT 1081
 Qy |||||
 1081 GTGACTTTGTATATTCATGACCAACAAATTTGAAACAGCCCTTACATCCAGTTTAGGG 1140
 Db |||||
 1082 GTGACTTTGTATATTCATGACCAACAAATTTGAAACAGCCCTTACATCCAGTTTAGGG 1141
 Qy |||||
 1141 ATCTTGATCAATGGGCAAAACCCAAATTTGGAAAAATTTCTGMAAAGAAAGAACTGTTCAG 1200
 Db |||||
 1142 ATCTTGATCAATGGGCAAAACCCAAATTTGGAAAAATTTCTGMAAAGAAAGAACTGTTCAG 1201
 Qy |||||
 1201 TTTGATGTCCAAAGTTGCGAATCTACTGTGACCCAGAAACAGAACACCGGGAGACAGA 1260
 Db |||||
 1202 TTTGATGTCCAAAGTTGCGAATCTACTGTGACCCAGAAACAGAACACCGGGAGACAGA 1261

1261 TGTGAGATTCCTGGATTTGCTTAAATGGTCCAGTGATGATAGGTTCAACTCCAGTCCC 1320
 Db |||||
 1262 TGTGAGATTCCTGGATTTGCTTAAATGGTCCAGTGATGATAGGTTCAACTCCAGTCCC 1321
 Qy |||||
 1321 TGTATTTGTCTCTCGGGAAACCCAGGACTTCAAGGCCCCCAAGGTGACCTTGACTGCCT 1380
 Db |||||
 1322 TGTATTTGTCTCTCGGGAAACCCAGGACTTCAAGGCCCCCAAGGTGACCTTGACTGCCT 1381
 Qy |||||
 1381 GGGAAACCTCGCTACCTGGACAACTGGTCAAGATGGTAAAGCCT----- 1425
 Db |||||
 1382 GGGAAACCTCGCTACCTGGACAACTGGTCAAGATGGTAAAGCCTGTGAGTACTGAAAGC 1441
 Qy |||||
 1426 -----GGATATCAGGAAATTCAGGGACACCAAGGT 1455
 Db |||||
 1442 TTAGTCATCTCCGGTATATCTGGGATACAGGATATCAGGAAATTCAGGGACACCAAGGT 1501
 Qy |||||
 1456 GTTCCAGGATCTCAGGAATACAGGAGCTCGAGGACTACCGAGTTACAAAGGAGAACCA 1515
 Db |||||
 1502 GTTCCAGGATCTCAGGAATACAGGAGCTCGAGGACTACCGAGTTACAAAGGAGAACCA 1561
 Qy |||||
 1516 GGGGACATGTCGACAAAGGTGATCGTGAGACTTCTCTGGTTTCTCGGCTTCATGGCATG 1575
 Db |||||
 1562 GGGGACATGTCGACAAAGGTGATCGTGAGACTTCTCTGGTTTCTCGGCTTCATGGCATG 1621
 Qy |||||
 1576 CCAGGATCAAAGGTGAATGGTGCCAAAGGAGACAAAGGATCACTGGATTTTATGGC 1635
 Db |||||
 1622 CCAGGATCAAAGGTGAATGGTGCCAAAGGAGACAAAGGATCACTGGATTTTATGGC 1681
 Qy |||||
 1636 AAAAAAGGTGCAAAAGGTGAAA 1657
 Db |||||
 1682 AAAAAAGGAAAGTATAGATCACA 1703
 Qy |||||
 1703 AAAAAAGGAAAGTATAGATCACA 1703
 Db |||||
 RESULT 3
 AAK53227
 ID AAK53227 standard; cDNA; 1578 BP.
 AC AAK53227;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 2756.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WC200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QH, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
 XX
 WP1: 2001-476283/51.
 DR P-PSDB; AAK80094.

different human tissues, useful in recombinant DNA methodologies -
Claim 1; Page 201-202; 1095pp; English.

This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence encodes a polypeptide described in the disclosure of the invention.

Sequence 4160 BP; 1271 A; 795 C; 1026 G; 1068 T; 0 other;

Query Match	99.1%	Score	2839.6	DB	23	Length	4160
1st Local Similarity	99.5%	Pred.	No. 0				
atches 286	Conservative	0	Mismatches	4	Indels	9	Gaps
1	ATGGCTCACTATATTAACATTTCTCTCGATGGTTTTGGTGCCTGCTCTTCCAGAAATCTGTG	60					
203	ATGGCTCACTATATTAACATTTCTCTCGATGGTTTTGGTGCCTGCTCTTCCAGAAATCTGTG	262					
61	TTAGCTGAAGATGGGAAGTAAAGATCAAGATTGTGCTGCTCCGACAGATTAGTTTTC	120					
263	TTAGCTGAAGATGGGAAGTAAAGATCAAGATTGTGCTGCTCCGACAGATTAGTTTTC	322					
121	ATCTTAGATGGCTCTTATAGTGTGGCCCGACAAAACCTTTGAAATAGTGAAGAATGGCTT	180					
323	ATCTTAGATGGCTCTTATAGTGTGGCCCGACAAAACCTTTGAAATAGTGAAGAATGGCTT	382					
181	GTCAATATCACAAAAAACCTTTGACATAGGCGCGAAGTTTATTCAGAGTTGGAGTGTCTCAA	240					
383	GTCAATATCACAAAAAACCTTTGACATAGGCGCGAAGTTTATTCAGAGTTGGAGTGTCTCAA	442					
241	TATAGTGACTACCTGTGCTGGAGATTCTCTCGGAAGCTATGATTCAGAGAAACATTTCG	300					
443	TATAGTGACTACCTGTGCTGGAGATTCTCTCGGAAGCTATGATTCAGAGAAACATTTCG	502					
301	ACGGCAGCAGTGGAAATCCATCTACTTCTAGGAGGAAACACAAAGACACGGGAAGGCCATC	360					
503	ACGGCAGCAGTGGAAATCCATCTACTTCTAGGAGGAAACACAAAGACACGGGAAGGCCATC	562					
361	CAGTTTGGCTCGATTACCTTTTGGCCAGTCTCTACGATTTCTGACTAAGATAGCAGTG	420					
563	CAGTTTGGCTCGATTACCTTTTGGCCAGTCTCTACGATTTCTGACTAAGATAGCAGTG	622					
421	GTACTTTACCGATGGCAATCCCAAGATGACGTCAAGGATGCAGCTCAAGCAGCAAGAT	480					
623	GTACTTTACCGATGGCAATCCCAAGATGACGTCAAGGATGCAGCTCAAGCAGCAAGAT	682					
481	AGTAAGATAACATTATTGTGCTATTGTGTTGGTTTCAGAAACAGAGATGCCGAATCTAGA	540					
683	AGTAAGATAACATTATTGTGCTATTGTGTTGGTTTCAGAAACAGAGATGCCGAATCTAGA	742					
541	GCTATTGCCAAACAAAGCTTCGCTACTATGTGTTTATGTGGAAGACTATATTGGCAATA	600					
743	GCTATTGCCAAACAAAGCTTCGCTACTATGTGTTTATGTGGAAGACTATATTGGCAATA	802					
601	TCCAAAATAGGGAAGTGTAGTGAAGCAGAACTTTGTGAAGATCTCTGCTCCACACAGA	660					
803	TCCAAAATAGGGAAGTGTAGTGAAGCAGAACTTTGTGAAGATCTCTGCTCCACACAGA	862					
661	ATTCCAGTGGCAGCTCGTGTATGAAGGGGATTTGATATTCTTTTAGGTTTAGATGTAAT	720					
863	ATTCCAGTGGCAGCTCGTGTATGAAGGGGATTTGATATTCTTTTAGGTTTAGATGTAAT	922					
721	AAAAAGGTTAAGAAAAGATATACAGCTTTTACAAAAAAGATATAAGAGTATGAAGTAAACA	780					
923	AAAAAGGTTAAGAAAAGATATACAGCTTTTACAAAAAAGATATAAGAGTATGAAGTAAACA	982					

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 18:17:32 ; Search time 715 Seconds
(without alignments)
10816.629 Million cell updates/sec

Title: US-09-996-611B-5
Perfect score: 2865
Sequence: 1 atgcttactatattacatt.....gaaaaggaccaaactattag 2865

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 252756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	N_Geneseq	19Jun03.*
1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*	
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*	
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*	
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*	
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*	
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*	
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*	
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*	
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*	
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*	
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*	
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*	
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*	
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*	
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*	
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*	
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*	
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*	
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*	
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*	
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*	
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*	
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*	
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*	
25:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2839.6	99.1	4160	23	ABX71234 Cell structure and
2	1589.2	55.5	1826	22	AAK52243 Human polynucleoti
3	1464.4	51.1	1578	22	AAK53227 Human polynucleoti
4	1464.4	51.1	1578	22	AAK53249 Human polynucleoti
5	1462.8	51.1	1691	22	AAK52265 Human polynucleoti
6	1282	44.7	2375	22	AAK52603 Human cDNA encodin
7	1282	44.7	2375	25	ABX73394 Human novel polynu
8	1062	37.1	2230	22	AAK52045 Human polynucleoti

ALIGNMENTS

RESULT 1

ABX71234
ID ABX71234 standard; cDNA; 4160 BP.

AC ABX71234;

DT 15-APR-2003 (first entry)

DE Cell structure and mobility-associated cDNA from clone DKFZphfbr2_2b5.

XX Human; gene; gene therapy; vaccine; disease treatment; detection; ss.

OS Homo sapiens.

PN NC200112659-A2.

XX 22-FEB-2001.

PF 18-AUG-2000; 2000WO-1B01496.

PR 18-AUG-1999; 99US-0149499.

PR 28-SEP-1999; 99US-0156503.

XX (GEHU-) GERMAN HUMAN GENOME PROJECT.

PI Wiemann S;

DR WP1; 2001-327840/34.

DR P-PSDB; ABU52683.

XX Nucleic acids having the sequences of clones isolated from libraries of

PT

Human polynucleoti
Human polynucleoti
Human polynucleoti
Human MDT encodin
Human MDT encodin
Human MDT encodin
Human diagnostic a
Human dthp polynu
EST clone AJ54. H
Human collagen XXI
Porcine collagen XI
Bovine alpha1(III)
Bovine alpha1(III)
Vector pAC3A1 cont
Collagen type III
Human Tumour Erdot
Breast cancer asso
DNA encoding novel
Human mRNA differe
Human EST-derived
Human EST-derived
Human mRNA differe
Human breast cance
Breast cancer-asso
Colon adenocarcino
Lung cancer relate
Human endometri.a
Lung cancer-associ
Human cDNA differe
IV collagen. alpha
DNA encoding novel
Splice variant ZAP
Shrimp white spot
Shrimp white spot
Type III procollag
DNA encoding novel

Db	1875	TGAACCCGCTTACCTGGACCCACAGGTTACATTGGGCCACAGGGTTCGTGGTCC	1934
Qy	2790	AGGCATCTGGACCCCATCACTATGTTTGTAGTGTAAATTGCCAGAAGAGATCCGTTCAGAAA	2849
Db	1935	TGGTTACTGGACCCCATCTACTTGTGTGATGGGTCAAGTCCGTGGCATTTCAACAACCTAA	1994
Qy	2850	AGGACCAAA	2858
Db	1995	AGGTCCAAA	2003

Search completed: October 2, 2003, 22:19:10
Job time : 10:51 secs


```

Db      2732 CTATGGAAATATGGCCGAGACGGTGGAGGAGGCCCCCAGGGCTGGCAGGAATCTCTGG 2791
QY      2757 AAAAAGCTGAATCAAGGCAACAGGCCCCCAGGCATCTCGACCCATC 2807
Db      2792 AGTGCTGGACCCCGGAGACCTCTGGGCTTCCCGGTTCTGTGAGCCAGC 2842

RESULT 15
AK115182
LOCUS   AK115182                2359 bp    mRNA    linear    INV 30-NOV-2002
DEFINITION   Ciona intestinalis cDNA, clone:cilv0.4e05, full insert sequence.
ACCESSION   AK115182
VERSION     AK115182.1    GI:23587129
KEYWORDS    FL1 CDNA.
SOURCE      Ciona intestinalis
ORGANISM    Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Clonidae; Ciona.

REFERENCE
1
AUTHORS    Satou,Y., Yamada,L., Mochizuki,Y., Takatori,N., Kawashima,T.,
            Sasaki,A., Hamaguchi,M., Awazu,S., Yagi,K., Sasakura,Y.,
            Nakayama,A., Ishikawa,H., Inaba,K. and Satoh,N.
            A cDNA resource from the basal chordate Ciona intestinalis
            Genesis 33 (4), 153-154 (2002)
            22191024
            12203911
            2 (bases 1 to 2359)
            Satou,Y. and Satoh,N.
            Direct Submission
            Submitted (04-OCT-2002) Nori Satoh, Kyoto University, Department of
            Zoology, Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            (E-mail:satohascidlan.zool.kyoto-u.ac.jp, Tel:81-75-753-4095,
            Fax:81-75-705-1113)
            Ciona intestinalis cDNA Project (URL:
            http://ghost.zool.kyoto-u.ac.jp/indexr1.html).
            Location/Qualifiers
            1..2359
            /organism="Ciona intestinalis"
            /mol_type="mRNA"
            /db_xref="taxon:7719"
            /clone="cilv0.4e05"

BASE COUNT  641 a 552 c 724 g 442 t

ORIGIN
Query Match      10.2%; Score 292.6; DB 3; Length 2359;
Best Local Similarity 50.3%; Pred. No. 3.5e-55;
Matches 789; Conservative 0; Mismatches 759; Indels 21; Gaps 2;

QY      1299 TGTAGTTCAATCCAGCTCCCTGTTATTTGCTCCCGGAAACACAGGACTTCAAGGCC 1358
Db      447 TGGAGCGCTACGGCTGTGAAAGAGTGGTTTCAATGGTGAAGTTGGTTTGTGGACC 506

QY      1359 CAAAGGTGACCTGGAGCTGCTGGGAACCTGGCTACCTCGACACCTGGTCAAGATGG 1418
Db      507 ACAGGTTATCCCGACCCAGGCTCAAGGATTAAGGATACAAGGTGAACCCGGTCTGATGG 566

QY      1419 TAAGCTTGATATCAGGGAATTCAGGGAACACAGGTGTTCCAGGATCTCCAGGAATACA 1478
Db      567 AAAACAAGTTCACAAAGGGTATCCCGTCTTGAGGAGAACAGGACGAGTGTCTCG 626

QY      1479 AGGAGCTCAGGACTACAGGTTACAAGAGAACACAGGCGAGATGGTGAACAGGGTGA 1538
Db      627 CGGTGAATCTGTGTCACGTGGGACAGGTTTACCGGGAGAGATGGGAGAAAGGGAGA 686

QY      1539 TCGTGACTTCTCTGGTTTCTCTGGGCTTCATGSCATGCCAGGATCAAGGGTGAATGG 1599
Db      687 AGTAGGCCCTGCAGGAGTACCTGGAGAAATGGTTTCACTCTGGTGAACAGGTTTACAGG 746

QY      1599 TGCCAAAGAGACAAAGGATCACCTGGATTTTATGGCAAAAAGGGTGCAAAAGGTCAAAA 1659
Db      747 AGCTGAAGTCTCTCTGGAGTCCAGGAGAGGGGAGCCAGGAGGAGAACTGGGGATCG 806

```

```

QY      1659 GGGAAATGCTGGCTTCCCTGGCTCTCTGGACCTGCTGGAGAACCAAGGAGCATCGAAA 1718
Db      807 AGGTCTCTTTGGAAAACCTGGAGGACCGGACTTCTTGGAGAACCGGGCTTACCCGGTAG 866

QY      1719 GGAATGAATTAATGGTAGTCCCGTTTCAAGGGAGAGCAGGATCCCTCTGGTCTCCGGG 1778
Db      867 AAATGGAGCTTATCCAGGAACCCAGGAAACCGGGAGACCAAGGAGAGTCAG 926

QY      1779 GCAGATGGACACAGCGGAGAGCTGAACTCCAGGATTTCTGTGAAACCGAGGATTAAT 1838
Db      927 ACCAAAGAGACCAAGAGTGCATAGAGCTCTCTGGAATACCTGGAGATATTGGACCTGA 986

QY      1839 GGGCAAAAGAGGAGAAATTTGGGCTCTCAGSACAGCAAGGAAAAAAGGAGCCCC----- 1891
Db      987 AGGACCAACTGGCAAAAGCTGGTGCACAAGGTATATCCAGGAGAGAAGTGGCGCCGTGGACC 1046

QY      1892 --CAGGATCCCTGGTTTATATGGAGAGCAATGGCTCACAGGCGACCTGGAAACACGGG 1949
Db      1047 AAAAGGAGAAAGTGGGACCAATTTGGCAAACTGGGTTTACAAAGTCCAGCTGGAGAGCAAG 1106

QY      1950 ATCTAAGGGAGCAAGAGTGAACCTGGAATTTCAAGGGATGCCCTGGGGCTTCTGGGCTCAA 2009
Db      1107 ACAGTGGAGATGCGAGACCATCAGAGAAATCTGSAACCACTTGGACCCGCTGGTAACAA 1166

QY      2010 GGGAGAACCAAGGACCAACCGGTTTCCCAAGAGAACCAAGGATACATGGSTTTACCCCGGAT 2069
Db      1167 GGGTAATTCAGGAGAACAGGTATTAGGGTCAACCTGGTGAACGTGGATTCCTCTGAAAT 1226

QY      2070 TCAAGGAAAAAGGGGACCAAGGAAATCAAGGTGAANAAGCTATTACAGGTCAAAAGG 2129
Db      1227 CAAAGGACACAAAGGAGCATCTGTTATGACTGGCGCAAGGGGAGAACCTGGTAAATGATGG 1286

QY      2130 AGAAATGGAGACAGGGAATTCAGGGCAACAGGGAATTCAGAGGCATCATGGTCAAAA 2189
Db      1287 AATGAGGACACAGAGGCAACCTCGACCAAGGAGAACTGTGGAACTCCAGSAGAACG 1346

QY      2190 AGGAGAGAGAGTGAAGAGGAGAACTGGTGTCCGAGGTGTCATTTGGATCAAAAGAGA 2249
Db      1347 TGGATCACAGGAAACACCGGTTGAAATGGGCAACGTGGCCAGGTGGACCAACCGGCAAT 1406

QY      2250 ATCTGGGTTGGATGGCTTGTATGGGCGCCGAGGCTCTAAGGGCAACCTGGGATCCAGG 2309
Db      1407 CATTTGACCAAGAGAGGACCCAGGGGCCAATCCGGGCCACAGGAGCAACAGGAGAAATCTGG 1466

QY      2310 TCTCTCAGGACCCCGAGGTTTGGATGGAGCCCGGAGAGAGATTTTCAGAACAAATTTAT 2369
Db      1467 GCCAAGAGGCGTCCCTGGATTACAAGCCCTCTCTGGCGCATGCCAAGTGACCAAAAGAT 1526

QY      2370 TCGACAAGTTTGCACAGATGTAATAGAGCCGAGCTACCAAGTCTTACTTCAGAGTGGAG 2429
Db      1527 TCTGAAATTTGGCGTAGGGTGGCTTCGAGATGAGATGAGAGGTATGATGGCAGCAACACAG 1586

QY      2430 AATTAGAAAATTGTGATCATTTGCTGCTCCCAACATGGCTCCCGGGTATTCTCTGGGCCACC 2489
Db      1587 ACCACAGGTATCATTTACTG-----TGAGCTCTCTGGAGACCGGACCTGT 1634

QY      2490 TGSTCCGATAGGCCACAGAGGTTCCAGAGATTTACCTGGTTTCCAGGAGAGAGATGGTGT 2549
Db      1635 TGGACCACTTGGACAGCAAGGACCAACAGCAATCTGGTCTTCTCTGTTAAATCTGAAT 1694

QY      2550 TCCTGGAATTAGTGGGTGCTCTCTGGAGCTCCAGGTGTACAGGATTTAAAGGSCCTACCAGG 2609
Db      1695 GCAAGGTGCAAGGGTGTCTCTGGAATGCTTGACCTACTGACCCCAAGGATTTGGTGG 1754

QY      2610 AAGAAATGGGCAAAAAGGAGGCAAGGTTTGGGTATCTCTGGAGAACAAAGGTCTCTCTGG 2669
Db      1755 TGATAAAGGTGACAGAGAGCGCCAGGAATTGGCAGCTTGGTCTCTCAGGACCTCTGG 1814

QY      2670 TCCCAAGGTCAGAGAGGCGCTCTCTGGAATTAAGCAAAAGAGGTCTCTCAGGAGACCCAGG 2729
Db      1815 TCTCTCAGGACCCCAAGGATCACCAGGTTTGGCAAGATGGTCTGTAAACGAGAACTGG 1874

QY      2730 TCTCCTCTGGCAAGATGGAGACCATGGAATACTTGAATCTCAAGAGGCAACAGGCCCCC 2789

```

REFERENCE 1 (bases 132 to 2927)
 MURAGAKI, Y., Kimura, T., Ninomiya, Y. and Olsen, B.R.
 The complete primary structure of two distinct forms of human alpha
 1 (IX) collagen chains
 Eur. J. Biochem. 192 (3), 703-708 (1990)
 JOURNAL 91006164
 MEDLINE 2209617
 PUBMED
 REFERENCE 2 (bases 1 to 2927)
 OLSEN, B.R.
 Direct Submission
 Submitted (28-JUN-1990) Olsen B.R., Dept. of Anatomy & Cellular
 Biology, Harvard Medical School, 220 Long Wood Ave., Boston, MA
 02115, USA
 COMMENT Data kindly reviewed (25-NOV-1990) by Olsen B.R.
 FEATURES Location/Qualifiers
 1..2927
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 132..2927
 /note="alpha (IX) collagen precursor"
 /codon_start=1
 /protein_id="CAA38276.1"
 /db_xref="GI:30084"
 /db_xref="SWISS-PROT:P20849"
 /translation="MKTWKIPVFVFVCSFLEPWASAAVKRRPRFPVNSNSNGNELC
 KIRIGODDLPGFDLISOFVDKASRRAIORVVGSSATQVAVKLGNNVDFRIPTNRL
 YPSGLPEYSLELTFRMTGSLTKKWNINQIODSGKBONGIKXINGOTQSVFYSYKGL
 DLSLQTAAPSNLSLSFDQSHKIMIGVERSSATLFDNCNRIESLPDKRPIDIDGFA
 VGLKADNPQVSPFELQMLIHCDLPRPRETHLHELPAITPSCTDERGPGEQGP
 PGASGPPVPGIDGIDGRPKGPPGAPGEPKPGAPKPGTGGADGLTGPDGSP
 GSIGSGKQKGEVPGSRGPPGPGIPGPPGPGTAGLPGELGRVPGVDPGRGPPGP
 PGPGRPTIGPHDGLPLCPNAPPPRSPGPGPLPMRCHKAKGEIIEPGRQGHKGE
 GQOGLVGEVGAQGPAGGLRGTICLVGDKEKGARGLDGEPGPGGLPGAPGDGORG
 PGEAGPGRGAEGARGIPGLPGKGTGCLPGVDGRDGIIPMGPTGEPKPGPPGP
 AGLQLPVGPIPGAGVAGEKSTGAPKPGQMNKSPQGGPPGPPGPPGPPGPPGPP
 GSRGLGVLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLG
 EREGAGERGELGIDLPKPGKSGNPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 QCATGLPGVQGPGRAPTDQHIKQVCMRVIQEHFAEMAASLKRPSDQATGLPGRPP
 GPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
 IGLGDPGPASVKNRGRGERPPGLAGIPGVPPPPGPPGLPGFCEPASCTMQLVSEH
 LKNGTLRLTLAAWLISA"
 132..200
 /note="signal peptide"
 201..2924
 /product="mature alpha1 (IX) collagen"
 BASE COUNT 697 a 724 c 911 g 595 t
 ORIGIN

1617 ATCACTCGATTTTATGCAAAAGGGTGCAAAAGGTGAAAGGGGAATGCTGGCTTCCC 1676
 1664 AGAAGAGGTCCCAAGAGAGATAGAGGGCTGAGAGTCTAGAGGATTTCTGGTCTCCC 1723
 1677 TGGCTCTCCCTGACCTGCTGGAGAACCCAGGAAGACATGGAAGATGATTAATTAATGAGTAG 1736
 1724 TGGGCCCAAGAGAGACACGGGTGTCAGAGTGTGATGGCGGTGATGGATCCCTCGAAT 1783
 1737 TCCGCTTTCAAGGAGAGAGAGGATCCCTGCTGCTCCGGGGCAGAGATGGAACACCGGG 1796
 1784 GCTTGAACAAAGAGGTGAACCCAGGAACCTGGGCTCTCTGGTATCAGAGATTTGCAAGG 1843
 1797 AGAGCTGGAATCCCGAGATTTCTTGGAAACCCAGGATTAATGGGCTCAAAAGAGGAGAAAT 1856
 1844 GTTACAGGTGTACCTGGAATTTCTGGTGCAAGGGTGTCTGCTGTAAGAGGTATACAC 1903
 1857 TGGGCTCTCAGACAGCAAGAAAGAAAGAGCCCGAGGATGCTGGTTTATTTATGGGAAG 1916
 1904 AGGTGCTCAGGAAAGCTGTGTAGATGGGAAATTCAGGCAACCCGGCCCAACAGGGGCC 1963
 1917 CAATGCTCACAGCCAGCTTGGAAACCCGGATCTAAGGGAAGCAAGCTGAACCTGG 1976
 1964 TCAGAGAGGTGGACCCCGAGGACCCAGGGCTTCTGGCACTAGAGGAAATTTAGG 2023
 1977 AATTCAAGGGATGCTTGGGCTTCTGGCTCAAGGAGAACCCAGAGCAACGGGTTCCTCC 2036
 2024 ACCAGTGGGATCCCGAGGCTTACCAGGTAAATCTGGGTCTCTGGGTAGCCCTGGCTCCC 2083
 2037 AGAGAACCCAGATACATGCGTTTACCCGGGATTCAGGAAAGAGGGGCAACAGGAA 2096
 2084 TGGCTTGGCTTGGGCTTGGGCTTCTGGGATTAAGAGTACAGGGGTGTAGTCTGGTGA 2143
 2097 TCAAGGTGAAAGGTATTTCAGGCTCAAAAGGAGGAAATGGAAGACAGGGAATTCACAG 2156
 2144 ACCGGTCAAAGGTGAACAGGTGCTCTGGTGAACAGGTGAAGCAGGAGAAAGGG 2203
 2157 GCAACAGGAAATCAAGGCTATGCTGTCAAAGAGAGAGAGGTGAAAGAGGAGAAC 2216
 2204 GGAATCTGAGATATAGGATTAACCTGGCCCAAGGGATCTTCAGGTAACTCTGGGAAAC 2273
 2217 TGGTCTCCAGGTGCAATTTGATCAAAAGAGGAAATCTGGGTGATGCTTGTGGGCTC 2276
 2264 TGGCTTGAAGGGCTGAGGAGGCTGGGGGCTTCTGGAGTGGAGAGCAAGAGGAC 2323
 2277 CAGAGTCTCAAGGGCAACCTGGGATTCAGGTCTCAGGACCCCGAGGTGTTGATGG 2336
 2324 ACCTGACCCCGGGGTGTCAGGAGAACAGGTGCCACCGGCTCTCTGGTGTCCAGG 2383
 2337 GAAGCCCGAGAGAGCTTTTCAGAACAAATTTATTCGACAAATTTGACAGATGTAATAG 2396
 2394 CCTCCGGGTAGAGCACCAGACAGATCAGACATTAAGCAGGTTCATGAGAGTATACA 2443
 2397 AGCCAGCTACAGCTTACTTTCAGAGTGAAGAAATAGAAATTTGATCATTTGCTGTC 2456
 2444 AGAACATTTGCTGAGATGGCTGCCAGTCTTAAGGCTCCAGACTCAGCTGCAC 2497
 2457 CCACATGCTCCCGGGTATTCCTGGGCACTGCTGGTTCGATAGGCGGAGGCTCCAG 2516
 2498 -----TGGGCTTCTTGAAGGCTGGGCTCTCTGGTCTCCCGGCTCTCTGGAGAA 2551
 2517 AGGATTACCTGGTTTCCAGGAGAGATGCTGCTCTGGATTAAGTGGGTGCTCTGGACG 2576
 2552 TGGTTTCCAGGCGAGATGGGATTCGTGGCTTCCGGGCTTAAGGGTCCCTCTGGTGC 2611
 2577 TCAGGTGTCAAGGATTAAGAGGCTTACAGGAAGAAATGGGAAAGAGGAGGCAAGG 2636
 2612 TCTTGGTTTTCAGGGACCTTAAAGGTGACTTGGGAGAAAGGGGAGCGCTGGCTCCAGG 2671
 2637 GTTGGGTATCTCGAGAACAGGTCTCTCTGCTCCCGGCTCCAGAGGCGCTCTCTGG 2696
 2672 AGAGTCTCCACAGGTTCCTCCAGGAGCTATAGGTCTCCCAAGTCCAGGCGCTTCGAG 2731
 2697 AATAAGAAAGAGGTCTCTCCAGGAGACCCAGGTCTCTCTCGCAAGATGGAGACCATGG 2756

(M1) collagen contains domains COL-3 and COL-2 connected by NC-2, and LMW (M2) collagen contains COL-1 together with part of NC-1.

FEATURES

```

source
    1..3164
        /organism="Gallus gallus"
        /mol_type="mRNA"
        /db_xref="taxon:9031"
    <1..3164
        /product="H/L mRNA"
    <1..2219
        /note="HMW/LMW collagen subunit precursor"
        /codon_start=3
        /protein_id="AA48675.1"
        /db_xref="GI:211499"
        /translation="LMISVTEVTLFDICIKVETLNKPKKISVDGFSVLRLKKN
        PDISVFEVQWMPHICDLRLPQREGCGLPARIISQTVIERGLPGPPGPPGPPG
        GIDGIDIRGNGPPGPPGPDGADKAGSPGLPEPDAGSLTGPDSFGATGPKQKG
        EPGPPGRLGKGLGPPGAGAGLPGVVRAGPPDPKRRPPGPPGPPGPPGPPG
        GLQDDPLCPNACPPRPGHAGLMKQKSGESGLGKQYKGEQGGPPGPRGTI
        AQPLGIPGIRGITITGPKNGKARGLDGGPGPLPGAPGGQGRGPGVEEGPKGE
        RPPQTRGILGPKGSGPLGVDRGCIIPMPKAKGEPKPGTGDGTGPPGPPG
        GSPGMKGIIPKNGRPPGVLGMSKPGEGPEGAGPTGPRGPPGSRGEGPAG
        PPGLPKGWPGDIDGLPGLPDPGPPGLPGKDRGSAGERPGKGEQAGSGDAGEKD
        LQDMGLPKAGSGNPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
        GPAGKEPTDQHIKQVMRMEQSLAASLRPEFGAPGLPGRPPGPPGPPG
        FPGQLGPRGLKGPGBIQRKPKGEAGERGERFGRGVKGLPGRGLPGRGPR?
        SYREGRDGVRRPGVAGPGIPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
        847 a 743 c 898 g 676 t
        BASE COUNT
        ORIGIN 121 bp upstream of HinfI site.

Query Match      10.6%; Score 303; DB 5; Length 3164;
Best Local Similarity 51.2%; Pred. No. 1.6e-57;
Matches 764; Conservative 0; Mismatches 715; Indels 12; Gaps 2;

QY 1317 TCCCTGTATTCTCTCCGGGAAACACGAGCTTCAAGGCCCAAGGTCGACCTGGACT 1376
DB 683 TCCCAATGCTTGTCCACCTGGCGCCGACGACATCTGCTGCTTAATGGGAATGAAGGACA 742

QY 1377 GCCTGGGAACCTCGCTACCTCGGACAACTTGGTCAAGATGGTAAGCCTGGATATCAGG 1436
DB 743 GAAGGCTCAAAAGGAGAACTCGGCGAATAAGTAAAGGCTTATAAGGCTGAAGAGG 802

QY 1437 AATTGACGGACACAGGTGTTCCAGGATCTCAGGAATACAGAGGCTCAGAGCTACC 1496
DB 803 GGGTCAAGTCCCAATGTTGAAGTTGGAGCTCAAGGCCCTTAGGCATCCAGAGTATCAG 862

QY 1497 AGGTTACAAAGGAGAACACGCGCGAGATGGTGACAGGCTGATCGTGACTTCTCGTTT 1556
DB 863 AGGTATACCGGATTAACAGGACCTAAGTAAACAAAGAGCTCTGTGGCTTGATGGCG 922

QY 1557 TCTTGGGCTTATGGCATGCCAGGATCAAGGGTGAATGGGTGCCAAGGAGACAAAG 1616
DB 923 TCTTGGTCCCAAGGTCTTCTGTGTGACCTCGGGGTCAAGGACAGAGAGTCCAGTGG 982

QY 1617 ATCACTGTGATTTATGGCAAAAGGGTGCAAAAGGTGAAGAGGGAATGCTGGTCTCC 1676
DB 983 AGAAGAGGTCCAAAGGGAGAAAGGTTCTCAAGGACAGAGGAATAAAGCTCTCTCC 1042

QY 1677 TGGCTCTCTGACCTGTGAGAACACAGGAGACATGGAAGATGGAATTAATGGGTAG 1736
DB 1043 TGGGCTTAAAGGAGAGTCTGGCTTGCCAGGAGTTGATGGCCGGGAAGGTATCCCTGGAA 1122

QY 1737 TCCCGTTTCAGGGAGAGCAGGATCCCTGCTCGGGGAGGATGGAACACGGG 1796
DB 1103 GCGCCGAGCAAAAGGGTGAACAGGAAACCTGGTATCCAGGTGATCAGGACCTCCCGG 1162

QY 1797 AGAGCTCGGAATCCCAAGGATTTCTTGGAAACCGAGATTAATGGGCCAAAGGAGGAAT 1856
DB 1163 ATTGCAAGTTTACCAAGGCTCTCCAGGTATGAAGGATTCCTGGCCCAAGGGTAATAG 1222

QY 1857 TGGGCTCCAGGACAGCAAGGAAAAAAGAGCCCGAGGATGCTGGTTTAATGGGAAG 1916
DB 1223 AGGTCTCTCGAGTCCCGGTTTGTATGGGAATTTCTGCAAAACCGGGTGAACAGGAGC 1282

```

RESULT 14

```

HSCOL9AL
LOCUS      Human mRNA for alpha1(IX) collagen (long form).
DEFINITION X54412
ACCESSION X54412.1 GI:30083
VERSION    collagen; collagen type IX.
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

HSCOL9AL      2927 bp      mRNA      linear      PRI 23-MAR-1995

```

```

QY 1917 CAATGCTCACAGGCGAGCCTGGAAACACCGGGATCTAAGGGAACAAAGGTGAACCTGG 1976
DB 1283 AGAGGGGAGAGAGGTCCNACGGGACCCCGGGGACACCGGGTAGCAGAGGAGAGCC--- 1339
QY 1977 AATTCAAGGATGCTCTGGGGCTTCTGGGCTCAAGGGAGAAACCCAGGAGCAACGGGTCC 2036
DB 1340 -----AGTCTGTCAGAGTCTCTCCAGGTTTACCAGGGAATGGGGTCCCAAGGATGACAT 1393
QY 2037 AGGAGAACAGAGTACATGGGTTTACCCGGGATTTCAAGGAAAAAAGGGGAGCAAGGAAA 2096
DB 1394 CGGATTCCTGAGATGCGCAGCTCTCCAGGCTTACTGCTGTAAGGGTGACCCGGGTTC 1453
QY 2097 TCAAGGTGAAAAGAGTATTCAAGGCTCAAAAGGAGAAATAAGGAAACAGAGGAATTTCCAGG 2156
DB 1454 AGCGGTGAACAGGACCTAAGGGTGACAGGTGCACTGATCAGAGGAGATGTCAGG 1513
QY 2157 GCAACAGGAAATTCAGGCTCATGTTGTCAAAAGAGAGAGAGGTGAAAAGGAGAAACC 2216
DB 1514 AGAAAAGGAGTACTTAGGTGATATGGGAATACCAGGTGCAAAAGGATCTGTTGGTAAATCC 1573
QY 2217 TGGTGTCCAGAGTGCATTTGGATCAAAAGGAGAAATCTGGGGTGGATGCTTGATGGGGC 2276
DB 1574 TGGAGACCCCGGTTCCGTTGGGCGGAGGAGAGTGGGGACCTGCTGGGATGGAAGGGCC 1633
QY 2277 CGCAGTCTCTAAGGGGCAACCTGGGATCCAGGTCTCTCAGGACCCCGCAGGTTTGGATGG 2336
DB 1634 ACGAGGTGCACCTGGACACAGAGGTTGCAGAGTGAACAGGTGCTCCAGGTCTGCTGG 1693
QY 2337 GAAGCCCGAAGAGAGTTCAGAAACATTTATTCGACAAAGTTTGCACAGATGTAATAG 2396
DB 1694 CAGCCAAAGTCCAGCTGGAAGAAACCAACTGATCAGCACATTAAAGCAAGTCTGCATGAG 1753
QY 2397 AGCCAGCTACCAAGTCTTACTTCAGAGTGGAAAGAAATAGAAATCTGTGATCATTTGCTGT 2456
DB 1754 AGTCATGCAAGAACAACTCTCAGTGCAGCCAGTCTTAGCGCGCCAGATTTTGGTGGC 1813
QY 2457 CCAACATGGCTCCCGGGTATTCCTGGGCCACTGTCGGATAGAGCCAGAGGGTCCAG 2515
DB 1814 TCCA---GGTCTCTCGGCGGACCTGGACCTCAGGTGCTCCAGGGCTCTCTGGGAAA 1870
QY 2517 AGATTACTGTTTCCAGAGAGAGATGGTGTCTCTGATAGTGGGTGCTCCCTGGAGC 2576
DB 1871 TGGTTTCCAGAGACAGCTTGGGCGCGGTGGCTGCTGAGGCTTAAAGGTCTCTCTGGTGA 1930
QY 2577 TCAGGTGTCAAGAGTATTAAGGCTTACAGAGAAATAGGGGAAAGAGGAGCAAGG 2636
DB 1931 GATCGCTCGTAAGGTCCTCAAGAGTGAAGCAGAGAGAGAGGAGAGGATTTCCAGG 1990
QY 2637 GTTTGGGTATCTGGAGAACAGAGTCTCTGCTGTCAGGATCCAGAGGGCCCTCTGG 2696
DB 1991 CAGAGAGTGAAGGTCTCCCGGACCGAGAGGTCTCCAGGTGAACACGAGCAACCCAG 2050
QY 2697 AATAAGCAAAAGAGTCTCCAGGAGACCCAGGCTCTCCCTGGCAAGATGGAGACCATGG 2756
DB 2051 CTATGCGAGGAGAGGCTGATGTTGATAGGATCTCCCTGCGGTGGCGGTGAGCCTGG 2110
QY 2757 AAAACCTGGAATCCCAAGGCAACCCAGGCGCCCGAGGATCTGCGACCCATC 2807
DB 2111 GATTCCTGCTCTGCTGCGGCTCCCGGCGCTCTCGGTACTGCGAGGCGCTC 2161

```

FEATURES	F#	Key	Location/Qualifiers	Location/Qualifiers
source	1.	.447	/organism="Zea mays"	
			/mol_type="genomic DNA"	
			/db_xref="taxon:4577"	
BASE COUNT	141	a	101	g
ORIGIN				
		Query Match	12.3%	Score 353.4; DB 6; Length 447;
		Best Local Similarity	98.3%	Pred. No. 4.8e-69;
		Matches 357;	Conservative 0;	Mismatches 6; Indels 0; Gaps 0;
Qy	1	ATGGCTCACTATATTACATTTCTCGATGGTTTGGTGCCTCTTCTCAGAAATCTGTG	60	
Db	85	ATGGCTCACTATATTACATTTCTCGATGGTTTGGTGCCTCTTCTCAGAAATCTGTG	144	
Qy	61	TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTGCTACTGCTCCACAGATTTAGTTTTC	120	
Db	145	TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTGCTACTGCTCCGACAGATTTAGTTTTC	204	
Qy	121	ATCTTAGATGGCTCTTATAGTTGTGGCCGAGAAACTTTGAAATAGTGAAGAAGTGGCTT	180	
Db	205	ATCTTTAGATGGCTCTAAATTGTGTGGCCGAGAAACTTTGAAATAGTGAAGAAGTGGCTT	264	
Qy	181	GTCATATACAAAAAACTTTGCATAGGGCCGAGCTTTATTCAGATTTGGAGTGGTTCAA	240	
Db	265	GTCATATACAAAAAACTTTGCATAGGGCCGAGCTTTATTCAGATTTGGAGTGGTTCAA	324	
Qy	241	TATAGTGACTACCTGTGCTGGAGATTCCTCTCGGAGCTATGATCAGGAGACATTTG	300	
Db	325	TATAGTGACTACCTGTGCTGGAGATTCCTCTCGGAGCTATGATCAGGAGACATTTG	384	
Qy	301	ACGGCAGCAGTGGAAATCCATACCTCTAGTAGGAGSAACAAAGACAGGGGAAGCCATC	360	
Db	385	ACGGCAGCAGTGGAAATCCATACCTCTAGTAGGAGSAACAAAGACAGGGGAAGCCATC	444	
Qy	361	CAG 363		
Db	445	GAG 447		
RESULT 13				

LOCUS	CHICKEN CHONDROCYTE SPECIFIC SHORT COLLAGEN, encoding HMW/LMW precursor.	3104 bp	mmNA	linear	Nov 28-Apr 1990
ACCESSION	K01702				
VERSION	GI:211498				
KEYWORDS	collagen; matrix protein.				
SOURCE	Gallus gallus (Chicken)				
ORGANISM	Gallus gallus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus				
REFERENCE	1. (bases 1 to 3164)				
AUTHORS	Ninomiya, Y. and Olsen, B.R.				
TITLE	Synthesis and characterization of cDNA encoding a cartilage-specific short collagen				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 81 (10), 3014-3018 (1984)				
MEDLINE	84221895				
PUBLISHED	6328487				
COMMENT	Original source text: Chicken embryo (17 days) sternal cartilage and calvarial bone, cDNA to mRNA, clone pYN1738. This mRNA, which codes for a collagenous polypeptide, has several unusual characteristics. First, the total length of the molecule is only about half that of pro-alpha-1(I) chains. Second, the molecule contains three collagenous domains connected by short, noncollagenous peptides. Third, several cysteinyl residues are located within the noncollagenous peptides. In the sites table NC denotes a noncollagenous region, COL denotes a collagenous region. [1] reports that this mRNA encodes the polypeptide precursor of one of the subunits of HMW and LMW collagens. They suggest that HMW				

[illegible]

```

RESULT 10
HS708F5/c
LOCUS
DEFINITION
HS708F5 126501 bp DNA linear PRI 04-MAR-2003
Human DNA sequence from clone RP4-708F5 on chromosome 6p11.2-12.3,
complete sequence.
AL031782
AL031782.1 GI:4164360
HTG.
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 126501)
Griffiths, C.
Direct Submission
Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerv@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jan 19, 1999 this sequence version replaced gi:3947843.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerv@sanger.ac.uk
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

FEATURES
source
Location/Qualifiers
1. .126501
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db xref="taxon:9606"

```

BASE COUNT	44:77 a	23605 c	21902 g	36813 t
ORIGIN	/chromosome="6"			
	/map="p11.2-12.3"			
	/clone="RP4-708F5"			
	/clone.lib="RPC1-4"			
Query Match.	19.3%	Score 553.6;	DB 9;	Length 126501;
Best Local Similarity	95.2%	Pred. No. 4.9e-114;		
Matches 571;	Conservative 0;	Mismatches 29;	Indels 0;	Gaps 0;
QY	41	TGCCTTCTCAGAAATCTCTGTGTAGCTGAAGATGGGGAAGTAGATCAAGATCAAGTGTGCTACTG	100	
DB	50175	TGCCTTTTATATAAAAGATGTGCTATGATGGTTTCCTTTTCACCTTAGGTTGTGCTACTG	50:16	
QY	101	CTCCGACAGATTAGTGTTCATCTTAGATGGCTCTTATAGTGTGGCCAGAAACTTTTG	160	
DB	50115	CTCCGACAGATTAGTGTTCATCTTAGATGGCTCTTATAGTGTGGCCAGAAACTTTG	50056	
QY	161	AAATAGTGAANAAGTGGCTTGTCAATATCAAAAAAACTTTGACATAGGGCGGAAGTTTA	220	
DB	50055	AAATAGTGAANAAGTGGCTTGTCAATATCAAAAAAACTTTGACATAGGGCGGAAGTTTA	49996	
QY	221	TTCAAGTTGGAGTGGTTCATATAGTGACTACCCCTGTGCTGGAGATTCCTCTCGGAAGCT	280	
DB	49995	TTCAAGTTGGAGTGGTTCATATAGTGACTACCCCTGTGCTGGAGATTCCTCTCGGAAGCT	49936	
QY	281	ATGATTTCAGGAGAAATTTGACGGCAGCGAGTGGAAATCCATACTTACTTAGAGAGAAACA	340	
DB	49935	ATGATTTCAGGAGAAATTTGACGGCAGCGAGTGGAAATCCATACTTACTTAGAGAGAAACA	49876	
QY	341	CAAGACAGGGAAGGCCATCCAGTTTGGCTCGATACCTTTTCCCAAGTCCCTCAGCAT	400	
DB	49875	CAAGACAGGGAAGGCCATCCAGTTTGGCTCGATACCTTTTCCCAAGTCCCTCAGCAT	49816	
QY	401	TTCTGACTAAGATAGCAGTGGTACTTACGGATGGCAATCCCAAGATCACGTCAGAGATG	460	
DB	49815	TTCTGACTAAGATAGCAGTGGTACTTACGGATGGCAATCCCAAGATCACGTCAGAGATG	49756	
QY	461	CAGCTCAGACGACAGAGATAGTAGATAACATTTATTTGCTATTGGTGTGGTTCAGAAA	520	
DB	49755	CAGCTCAGACGACAGAGATAGTAGATAACATTTATTTGCTATTGGTGTGGTTCAGAAA	49696	
QY	521	CAGAAGATGCCGAACCTTAGAGCTATTGCCAACAGCCCTTCGCTACTTATGTGTTTATG	580	
DB	49695	CAGAAGATGCCGAACCTTAGAGCTATTGCCAACAGCCCTTCGCTACTTATGTGTTTATG	49636	
QY	581	TGGAGACTATATTCCATATCCAAATAAGGGAAGTGAATAGAGAGAAACTTTGTGAAG	640	
DB	49635	TGGAGACTATATTCCATATCCAAATAAGGGAAGTGAATAGAGAGAAACTTTGTGAAG	49576	
RESULT 11				
AX721209	1414 bp mRNA linear PAT 07-MAY-2003			
LOCUS	Sequence 169 from Patent WO0220754.			
DEFINITION	AX721209			
ACCESSION	AX721209.1 GI:30422045			
VERSION				
KEYWORDS	Homo sapiens (human)			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1			
AUTHORS	Lincoln,S.E., Altus,C.M., Dufour,G.E., Chalup,M.S., Hillman,J.L., Jones,A., Yu,C.Y., Wright,R.J., Gietzen,D., Liu, Toomy,F., Yap,P., Dahi,C.R., Moriyama,M., Bradley,D., Rohatgi,S., Harris,B., Rossberry,A.M., Gerstin,E.H., Peralta,C.H., David,M., Panzer,S., Flores,V., Daffo,A., Marwaha,R., Chen,A., Chang,S.C. and Inmar,R.R.			
TITLE	Molecules for diagnostics and therapeutics			
JOURNAL	Patent: WO 0220754-A 169 14-MAR-2002;			
FEATURES	Incyte Genomics, Inc. (US) Location/Qualifiers			

QY 961 TATTATTACAAAC 974
DB 1156 TATTATTACAAAC 1169

RESULT 8
AX552238
LOCUS AX552238 1115 bp DNA linear PAT 27-NOV-2002
DEFINITION Sequence 147 from Patent WO0162927.
ACCESSION AX552238
VERSION AX552238.1 GI:25896476
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Banville, S.C., Greenawalt, L.B., Lincoln, S.E., Stockdreher, T.K.,
Amshey, S.C., Chang, S.C., Chen, W., D'Sa, S.A., Dam, T.C., Liu, T.F.,
Rosen, B.H., Russo, F.D., Spiro, P.A., Bradley, D.L., Chen, A.,
Cohen, H.J., Daffo, A., Daniels, S.E., Dufour, G.E., Flores, V.,
Fong, W.T., Hodgson, D.M., Jackson, S., Jones, A.L., Panzer, S.,
Roseberry, A.M., Shah, P., Wright, R.J., Yap, P.E., Yu, J.Y.,
Bracher, S.R., Chalup, M.S., Dahl, C.R. and Hillman, J.L.
TITLE Polypeptides and corresponding polynucleotides for diagnostics and
therapeutics
JOURNAL Patent: WO 0162927-A 147 30-AUG-2001;
Incyte Genomics, Inc. (US)
FEATURES
source
1..1115
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: LG:120744.1:2000MAV19"
BASE COUNT 344 a 215 c 259 g 296 t 1 others
ORIGIN

Query Match 31.7%; Score 908; DB 6; Length 1115;
Best Local Similarity 99.8%; Pred. No. 1.5e-194;
Matches 919; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 ATGGCTCAGTATATACATTTCTCTGCATGGTTTGGTGGCTTCTTCCAGAAATCTGTG 60
DB 196 ATGGCTCAGTATATACATTTCTCTGCATGGTTTGGTGGCTTCTTCCAGAAATCTGTG 255
QY 61 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGTCTGATGCTGCTCCGACAGATTTAGTTTC 120
DB 256 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGTCTGATGCTGCTCCGACAGATTTAGTTTC 315
QY 121 ATCTTAGAGGCTCTTATAGTGTGGCCAGAAACTTTGAATAGTGAAAGTGGCTT 180
DB 316 ATCTTAGAGGCTCTTATAGTGTGGCCAGAAACTTTGAATAGTGAAAGTGGCTT 375
QY 181 GTCAATATCAAAAAAATTTGACATAGGCGCGAAGTTTATTCAAGTTGGAGTGGTTCAA 240
DB 376 GTCAATATCAAAAAAATTTGACATAGGCGCGAAGTTTATTCAAGTTGGAGTGGTTCAA 435
QY 241 TATAGTACTACCTGTGCTGAGATTCCTCTCGAAGCTATGATTCAGGAGACATTTG 300
DB 436 TATAGTACTACCTGTGCTGAGATTCCTCTCGAAGCTATGATTCAGGAGACATTTG 495
QY 301 ACGGAGCAGTGGATCCATCTCTACTTTAGGAGAAACACAAAGACAGGAGGCGCATC 360
DB 496 ACGGAGCAGTGGATCCATCTCTACTTTAGGAGAAACACAAAGACAGGAGGCGCATC 555
QY 361 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCTCAGATTCTGACTTAAGATAGCAGTG 420
DB 556 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCTCAGATTCTGACTTAAGATAGCAGTG 615
QY 421 GTACTTACGGATGGGAAATCCCAAGATGACGTCAAGGATGAGCTCAAGCAGCAGAGAT 480
DB 616 GTACTTACGGATGGGAAATCCCAAGATGACGTCAAGGATGAGCTCAAGCAGCAGAGAT 675

QY 481 AGTAAGATAACAATATTTGCTATTGGTGTGGTTTCAGAAACAGAGATGCGAACTTAGA 540
DB 676 AGTAAGATAACAATATTTGCTATTGGTGTGGTTTCAGAAACAGAGATGCGAACTTAGA 735
QY 541 GCTATTGCCAACAAAGCTTCGCTACTTATGTTGTTTATGTGGAAGACTATATTGCAATA 600
DB 736 GCTATTGCCAACAAAGCTTCGCTACTTATGTTGTTTATGTGGAAGACTATATTGCAATA 795
QY 601 TCCAAATTAAGGAGTGTGATGAAGCAGCAAACTTTGTGAAGATCTGCTCCCAACAGCA 660
DB 796 TCCAAATTAAGGAGTGTGATGAAGCAGCAAACTTTGTGAAGATCTGCTCCCAACAGCA 855
QY 661 ATTCCAGTGGCAGCTCGTGTATGAAGGGGATTTCATATTCCTTTAGGTTTAGATGTAAT 720
DB 856 ATTCCAGTGGCAGCTCGTGTATGAAGGGGATTTCATATTCCTTTAGGTTTAGATGTAAT 914
QY 721 AAAAAAGTTTAAGAAAGAAATACAGCTTTCACCAAAAAAGATAAAGAGATATGAAGTAAACA 780
DB 915 AAAAAAGTTTAAGAAAGAAATACAGCTTTCACCAAAAAAGATAAAGAGATATGAAGTAAACA 974
QY 781 TCAAAAGTTGATTATACAAACTCACAAGCAATGTTTTCCAGAAAGCTTTCCTCCATCA 840
DB 975 TCAAAAGTTGATTATACAAACTCACAAGCAATGTTTTCCAGAAAGCTTTCCTCCATCA 1034
QY 841 TATGTAATTTGTTCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGAATA 900
DB 1035 TATGTAATTTGTTCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGAATA 1094
QY 901 TTAACCTATTGATGGAGGCCA 921
DB 1095 TTAACCTATTGATGGAGGCCA 1115

RESULT 9
AF438327.2
WPCOMMENT
Sequence split into 4 fragments LOCUS AF438327 Accession AF438327
Fragment Name Begin End
AF438327.0 1 1:0000
AF438327.1 100001 210000
AF438327.2 200001 310000
AF438327.3 300001 381696
Continuation (3 of 4) of AF438327 from base 200001 (AF438327 Homo sapiens alpha 1 type
Query Match 19.3%; Score 553.6; DB 9; Length 110000;
Best Local Similarity 95.2%; Pred. No. 4.8e-114;
Matches 571; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 4: TCGCTTCTCAGAAATCTCTGTTAGTCTGAAGATGGGAAGTAAAGATCAAGTTTCTGTACTG 100
DB 27134 TCGCTTCTCAGAAATCTCTGTTAGTCTGAAGATGGGAAGTAAAGATCAAGTTTCTGTACTG 27193
QY 101 CTCGACAGATTTAGTTTTCATCTTAGATGGCTCTTATAGTGTGGCCCGAGAAACTTTG 160
DB 27194 CTCGACAGATTTAGTTTTCATCTTAGATGGCTCTTATAGTGTGGCCCGAGAAACTTTG 27253
QY 161 AATAGTGAAGAGTGGCTTCTCAATATCACAANAATTTTGACATAGGCGCGAAGTTTA 220
DB 27254 AATAGTGAAGAGTGGCTTCTCAATATCACAANAATTTTGACATAGGCGCGAAGTTTA 27313
QY 221 TTCAAGTTGGAGTGTTCATATATAGTACTACCTGTGCTGGAGATTCCTCTCGAAAGCT 280
DB 27314 TTCAAGTTGGAGTGTTCATATATAGTACTACCTGTGCTGGAGATTCCTCTCGAAAGCT 27373
QY 281 ATGATTAGGAGAAATTTGACGAGCAGTGGGAATCCATCTCTACTTAGGAGGAAACA 340
DB 27374 ATGATTAGGAGAAATTTGACGAGCAGTGGGAATCCATCTCTACTTAGGAGGAAACA 27433
QY 341 CAAACAGAGGAGGCCATCCAGTTTGGCTCGATTACCTTTTGGCAAGTCTCTCAGAT 400
DB 27434 CAAACAGAGGAGGCCATCCAGTTTGGCTCGATTACCTTTTGGCAAGTCTCTCAGAT 27493
QY 401 TTCTGACTAAGATAGCAGTGTACTTACGGATGGCAAAATCCCAAGATGACGTCAAGGATG 460

TITLE
JOURNAL

COMMENT
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kanetani, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction; Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing; RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; Clone selection for full insert sequencing; HRI and
RAB; annotation; HRI and RAB.

FEATURES
source
1..2295
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NTONG2007034"
/tissue_type="tongue"
/clone_lib="NTONG2"
/note="cloning vector: pME18SFL3"

BASE COUNT
683 a 427 c 565 g 620 t

ORIGIN
Query Match 36.6%; Score 1048; DB 9; Length 2295;
Best Local Similarity 99.8%; Pred. No. 3.4e-226;
Matches 1060; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1804 GGAATCCCAAGGATTTCTCGAACCAGGATTAATGGGCCCAAAAGGGAGAAATTTGGGCT 1863
DB 169 GGAATCCCAAGGATTTCTCGAACCAGGATTAATGGGCCCAAAAGGGAGAAATTTGGGCT 228
QY 1864 CCAGGACAGCAAGGAAAAAAGAGCCCGCAGGATGCTCGTTTAATGGGAAGCAATGGC 1923
DB 229 CCAGGACAGCAAGGAAAAAAGAGCCCGCAGGATGCTCGTTTAATGGGAAGCAATGGC 288
QY 1924 TCACGAGCCAGCTGGACACCGGGATCTAAGGAAGCAAGGATGAACCTGGATTCAA 1983
DB 289 TCACGAGCCAGCTGGACACCGGGATCTAAGGAAGCAAGGATGAACCTGGATTCAA 345
QY 1984 GGGATCCCTGGGCTCTGGGCTCAAGGGAGAACCCAGGACCAAGGGTTCCCGCAGGAGAA 2043
DB 349 GGGATCCCTGGGCTCTGGGCTCAAGGGAGAACCCAGGACCAAGGGTTCCCGCAGGAGAA 408
QY 2044 CCAGGATACATGGTTTACCGGGATTCAGGAAAAAAGGGGACCAAGGAATCAAGGT 2103
DB 409 CCAGGATACATGGTTTACCGGGATTCAGGAAAAAAGGGGACCAAGGAATCAAGGT 468
QY 2104 GAAAAGGTATTTCAGGGTCAAAAGGGAGAAATGAAGACAGGGAAATTCAGGGCAACAG 2163
DB 469 GAAAAGGTATTTCAGGGTCAAAAGGGAGAAATGAAGACAGGGAAATTCAGGGCAACAG 528
QY 2164 GGAATTCAGGGCATCATGGTGCAGAAAGGAGAGAGGTGAAGGGAGAACCTGGTGTG 2223
DB 529 GGAATTCAGGGCATCATGGTGCAGAAAG--GAGAGAGGTGAAGGGAGAACCTGGTGTG 586
QY 2224 CCAGGTGCCATTCAGATCAAAAGGAGAAATCTGGGGTGGATGGCTTCATGGGGCCCGCAGGT 2283
DB 587 CCAGGTGCCATTCAGATCAAAAGGAGAAATCTGGGGTGGATGGCTTCATGGGGCCCGCAGGT 646
QY 2284 CTTAAGGGCAACCTGGGATCCAGGTCTCCAGGGACCCCGCAGGTTCGGATGGGAAGGCC 2343
DB 647 CTTAAGGGCAACCTGGGATCCAGGTCTCCAGGGACCCCGCAGGTTCGGATGGGAAGGCC 706
QY 2344 GGAAGAGTTTTTCAGAACAAATTTATTCAGCAAGTTTCACAGATGTAATGAAGCCAG 2403
DB 707 GGAAGAGTTTTTCAGAACAAATTTATTCAGCAAGTTTCACAGATGTAATGAAGCCAG 766
QY 2404 CTACCAAGTCTTACTTCAGAGTGGAGAAATTAGAAATTTGATCATTCGCTCTCCCAACAT 2463
DB 767 CTACCAAGTCTTACTTCAGAGTGGAGAAATTAGAAATTTGATCATTCGCTCTCCCAACAT 826
QY 2464 GGCTCCCGGGTATTCTCTGGGGCAACCTGGTCCGATAGAGCCCGCAGAGGTCCCGAGAGGATTA 2523

0
DB 327 GGCTCCCGGGGTATTCCTGGGCGACCTGGTCCGATAGCCCCAGAGGGTCCCGAGAGGATTA 886
QY 2524 CTTGTTTTCAGGAGAGATGGTGTCTCTGGATAGTGGTGTCTCTCGACGTCAGGT 2583
DB 887 CTTGTTTTCAGGAGAGATGGTGTCTCTGGATAGTGGTGTCTCTCGACGTCAGGT 946
QY 2584 GTCACAGGATTAAGAGGCTACCAAGGAAGAAATGGGAAAAAGGAGCCCAAGGCTTTGGG 2643
DB 947 GTCACAGGATTAAGAGGCTACCAAGGAAGAAATGGGAAAAAGGAGCCCAAGGCTTTGGG 1006
QY 2644 TATCTCGAGAGCAACAGGTCTCTCTGGTCCCGCAGGTCCAGAGGGCCCTCTCGAATTAAGC 2703
DB 1007 TATCTCGAGAGCAACAGGTCTCTCTGGTCCCGCAGGTCCAGAGGGCCCTCTCGAATTAAGC 1066
QY 2704 AAAGAGGTCTCTCGAGGAGACCCAGGTCTCTCTGGCAAGATGAGACCATGGAACCT 2763
DB 1067 AAAGAGGTCTCTCGAGGAGACCCAGGTCTCTCTGGCAAGATGAGACCATGGAACCT 1126
QY 2764 GGAATCCCAAGGGCAACAGGCCCCCAGGATCTCGGACCCCATCATCTATGTTTAGTGTA 2823
DB 1127 GGAATCCCAAGGGCAACAGGCCCCCAGGATCTCGGACCCCATCATCTATGTTTAGTGTA 1186
QY 2824 ATTGCAGAGAGATCCGTTTCAGAAAGGACCAAACTATTAG 2865
DB 1187 ATTGCAGAGAGATCCGTTTCAGAAAGGACCAAACTATTAG 1228

RESULT 6
LOCUS AX430410 1184 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 192 from Patent WO0240715.
ACCESSION AX430410
VERSION AX430410.1 GI:21655774
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Chalup M.S., Altus C.W., Lincoln S.E., Dufour G.E. and Jackson S.
TITLE Volucules for disease detection and treatment
JOURNAL Patent: WO 0240715-A 192 23-MAY-2002;
INCITE GENOMICS INC (US)
FEATURES
source
1..1184
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Incyle ID No: LJ:108819.1:2000SEP08"

BASE COUNT
368 a 227 c 272 g 317 t

ORIGIN
Query Match 34.0%; Score 974; DB 6; Length 1184;
Best Local Similarity 100.0%; Pred. No. 1.8e-209;
Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTCACTATATTACATTTCTCTGCATGGTGTGGTGTCTCTCAGAAATTCGTG 60
DB 211 ATGGCTCACTATATTACATTTCTCTGCATGGTGTGGTGTCTCTCAGAAATTCGTG 276
QY 61 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGCTGCTACTGCTCCGACAGATTTAGTTTC 120
DB 271 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGCTGCTACTGCTCCGACAGATTTAGTTTC 330
QY 121 ATCTTAGATGGCTCTTATAGTGTGGGCCCGAGAAACCTTGAATAGTGAAGAGTGGCTT 180
DB 331 ATCTTAGATGGCTCTTATAGTGTGGGCCCGAGAAACCTTGAATAGTGAAGAGTGGCTT 390
QY 181 GTCAATATACAAAAAATCTTTGACATAGGCGCCGAGTTTATCAAGTTGAGTGGTCAA 240
DB 391 GTCAATATACAAAAAATCTTTGACATAGGCGCCGAGTTTATCAAGTTGAGTGGTCAA 450


```
QY 1081 GTGACTTTGTTATTTATGATGACCAACAAATTTGAAACCAAGCCCTTACATCCAGTTTATAGG 1140
DB 1283 GTGACTTTGTTATTTATGATGACCAACAAATTTGAAACCAAGCCCTTACATCCAGTTTATAGG 1342
QY 1141 ATCTTGATCAATGGGCAACCAACCAATTTGAAATATTTCTGAAAGAGAAACCTGTTTCAG 1200
DB 1343 ATCTTGATCAATGGGCAACCAACCAATTTGAAATATTTCTGAAAGAGAAACCTGTTTCAG 1402
QY 1201 TTGTATGTCCTCAAAAGTTGGAATTTACTGTGTGACCAAGAAACAAACCCGGAGACAGCA 1260
DB 1403 TTTGATGTCCTCAAAAGTTGGAATTTACTGTGTGACCAAGAAACAAACCCGGAGACAGCA 1462
QY 1261 TCTGAGATTCCTGGATTT-----TGCCTTAATGTCCTCCAGTGATGTTCAACT 1311
DB 1463 TGTGAGATTCCTGGATTTAATGGAGATGCTTTAATGTTCCAGTGATGTTAGGTTCAACT 1522
QY 1312 CCAGCTCCCTGTATTTGCTCTCGGGAACCAACAGGACTTCAAGGCCCCCAAAAGGTGACCCCT 1371
DB 1523 CCAGCTCCCTGTATTTGCTCTCGGGAACCAACAGGACTTCAAGGCCCCCAAAAGGTGACCCCT 1582
QY 1372 GCACTGCTGGGACCTGGTACCTTGGACGACCTGGTCAAGATGGTAAAGCTGGATAT 1431
DB 1583 GCACTGCTGGGACCTGGTACCTTGGACGACCTGGTCAAGATGGTAAAGCTGGATAT 1492
QY 1432 CAGGGAATTCAGGGACACCAAGGTGTTCCAGGATCTCCAGGAATACAAAGGAGCTCGAGGA 1491
DB 1643 CAGGGAATTCAGGGACACCAAGGTGTTCCAGGATCTCCAGGAATACAAAGGAGCTCGAGGA 1702
QY 1492 CTACAGGTTACAAAGGAACCAAGGCGAGATGGTGACCAAGGTGATCGTGGATTCCT 1551
DB 1703 CTACAGGTTACAAAGGAACCAAGGCGAGATGGTGACCAAGGTGATCGTGGATTCCT 1762
QY 1552 GGTTCCTCGGCTTCATGGCATGCCAGGATCAAGGGTGAAATGGGTGTCGCAAGGAGAC 1611
DB 1763 GGTTCCTCGGCTTCATGGCATGCCAGGATCAAGGGTGAAATGGGTGTCGCAAGGAGAC 1822
QY 1612 AAAGATACCTGGATTTTATGGCAAAAGGGTGCAAAAGGTGAAAGGGGAATGCTGGC 1671
DB 1823 AAAGATACCTGGATTTTATGGCAAAAGGGTGCAAAAGGTGAAAGGGGAATGCTGGC 1892
QY 1672 TTCCCTGGCTCCCTGGACCTGCGAGAACCAAGGACATGGAAGAGATGGATTAATG 1731
DB 1883 TTCCCTGGCTCCCTGGACCTGCTGGAGAACCAAGGACATGGAAGAGATGGATTAATG 1942
QY 1732 GGTAGTCCGGTTTCAAGGGAGAGCAGGATCCCTGGTGTCTCCGGGCAAGATGGAAACA 1791
DB 1943 GGTAGTCCGGTTTCAAGGGAGAGCAGGATCCCTGGTGTCTCCGGGCAAGATGGAAACA 2002
QY 1792 CGGGGAGAGCTGGATCCAGGATTTCTGGGAACCGAGGATTAATGGGCCAAAGGGA 1851
DB 2003 CGGGGAGAGCTGGATCCAGGATTTCTGGGAACCGAGGATTAATGGGCCAAAGGGA 2062
QY 1852 GAAATTTGGCTCCAGGACAGCAAGGAAAAAGGAGCCCAAGGATGCTGTTTAAATG 1911
DB 2063 GAAATTTGGCTCCAGGACAGCAAGGAAAAAGGAGCCCAAGGATGCTGTTTAAATG 2122
QY 1912 GGAAGCAATGGCTCACAGGCCAGCTGGAAACCCGGATCTAAGGGAGCAAGGTGAA 1971
DB 2123 GGAAGCAATGGCTCACAGGCCAGCTGGAAACCCGGATCTAAGGGAGCAAGGTGAA 2182
QY 1972 CTTGGATTCAGGGATGCTGGGGCTTCTGGGCTCAAGGGAGAACCCAGGAGCAACGGGT 2031
DB 2183 CTTGGATTCAGGGATGCTGGGGCTTCTGGGCTCAAGGGAGAACCCAGGAGCAACGGGT 2242
QY 2032 TCCCGAGGAAACCAAGATACATGGTTTACCCGGATTTCAAGGAAAAAGGGGACAAA 2091
DB 2243 TCCCGAGGAAACCAAGATACATGGTTTACCCGGATTTCAAGGAAAAAGGGGACAAA 2302
QY 2092 GGAATTCAGGTCGAAAGGATTTACGGGTCAAGGGAGAAATGGAGACAGGGAAT 2151
DB 2303 GGAATTCAGGTCGAAAGGATTTACGGGTCAAGGGAGAAATGGAGACAGGGAAT 2362
```

```
QY 2152 CCAGGGCAACAGGGAATTCAGGCCATCATGGTCCAAAGAGAGAGAGAGAGTGAAGGGA 2211
DB 2363 CCAGGGCAACAGGGAATTCAGGCCATCATGGTCCAAAGAGAGAGAGAGTGAAGGGA 2422
QY 2212 GAACTCTGGTGTCCGAGGTGCATTTGGATCAAAAGAGAAATCTGGGGTGGATGGCTTGATG 2271
DB 2423 GAACTCTGGTGTCCGAGGTGCATTTGGATCAAAAGAGAGAAATCTGGGGTGGATGGCTTGATG 2482
QY 2272 GGGCCCGGAGGTCTTAAGGGGCAACCTGGGGATCCAGGTCTCAGGAGACCCCAAGGTTG 2331
DB 2483 GGGCCCGGAGGTCTTAAGGGGCAACCTGGGGATCCAGGTCTCAGGAGACCCCAAGGTTG 2542
QY 2332 GATGGGAAGCCCGGAAGAGAGTTTTCAGAACAAATTTATCGACAAGTTTGACAGATGTA 2391
DB 2543 GATGGGAAGCCCGGAAGAGAGTTTTCAGAACAAATTTATCGACAAGTTTTCAGAGATGTA 2602
QY 2392 ATAAGAGCCCACTACCACTCTTACTTTCAGAGTGGAAAGATTAAGAAATTTGATCATTCG 2451
DB 2603 ATAAGAGCCCACTACCACTCTTACTTTCAGAGTGGAAAGATTAAGAAATTTGATCATTCG 2662
QY 2452 CTGTCCTCAACATGGCTCCCGGGTATTCCTGGGCCACCTGGTCCGATAGGCCAGAGGGT 2511
DB 2663 CTGTCCTCAACATGGCTCCCGGGTATTCCTGGGCCACCTGGTCCGATAGGCCAGAGGGT 2722
QY 2512 CCCAGAGGATTTACCTGGTTTCCAGGAAGAGATGGTGTCTCTGATTTAGTGGGTGTCCT 2571
DB 2723 CCCAGAGGATTTACCTGGTTTCCAGGAAGAGATGGTGTCTCTGATTTAGTGGGTGTCCT 2782
QY 2572 GCACTGCCAGGTGTTCAGAGGATTAAGAGGCTTACAGGAAGAAATGGGAAAAAGGAGC 2631
DB 2783 GCACTGCCAGGTGTTCAGAGGATTAAGAGGCTTACAGGAAGAAATGGGAAAAAGGAGC 2842
QY 2632 CAAGGTTTGGGATTCCTGGAGAACCAAGGTCTCTGTCCTCCAGAGTCCAGAGGGCCCT 2691
DB 2843 CAAGGTTTGGGATTCCTGGAGAACCAAGGTCTCTGTCCTCCAGAGTCCAGAGGGCCCT 2902
QY 2692 CTTGGAATTAAGCAAAAGAGGTCTCTCCAGGAGACCAAGGTCTCTCTGTCCTCCAGAGTGGAGAC 2751
DB 2903 CTTGGAATTAAGCAAAAGAGGTCTCTCCAGGAGACCAAGGTCTCTCTGTCCTCCAGAGTGGAGAC 2962
QY 2752 CATGGAACCTCGGAATTCAGGGCAACCAAGGCTCCAGGCCCTCCAGGCATCTCGAGCCCACTA 2811
DB 2963 CATGGAACCTCGGAATTCAGGGCAACCAAGGCTCCAGGCCCTCCAGGCATCTCGAGCCCACTA 3022
QY 2812 TGTTCCTAGTGAATTTGCGAGAGAGATCCGTTTCAGAAAGAGACCAAACTATTAG 2865
DB 3023 TGTTCCTAGTGAATTTGCGAGAGAGATCCGTTTCAGAAAGAGACCAAACTATTAG 3076
```

RESULT 5
AK096444

LOCUS AK096444 2295 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ39125 fis, clone NTONG2007034.
ACCESSION AK096444
VERSION AK096444.1 GI:21755943
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1
Ogihara, A., Takahashi-Fujii, A., Tarase, T., Imose, N., Takeuchi, K.,
Arita, M., Musashino, K., Yuchi, H., Hara, H., Sugiyama, T., Irie, R.,
Ctsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Iseno, Y.,
Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yakashita, H.,
Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
Wagatsura, M., Murakawa, K., Kurehori, K., Sugiyama, A., Kawakami, S.,
Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and
Isegai, T.
NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2295)
AUTHORS Isegai, T. and Yamamoto, J.

Db 923 |||||TTAAAGGTTTAAAGAAAGATACAGCTTTTCCACCAAAAAGATATAAAGGATATGAAGTAAACA 982
Qy 781 TCAAAAGTTGATTTATTCAGAACTCAACAGCAATGTTTTCCAGAGAGTCTTCTCCATCA 840
Db 983 TCAAAAGTTGATTTATCAAGAACTCAACAGCAATGTTTTCCAGAGAGTCTTCTCCATCA 1042
Qy 841 TATGATTTTGTCTCTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGAGAA 900
Db 1043 TATGATTTTGTCTCTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGAGAA 1102
Qy 901 TTAACCTATTGATGGAAGGCCACAAATAGCAGTTACCTTAAATGGTGTGGACAAATCTTA 960
Db 1103 TTAACCTATTGATGGAAGGCCACAAATAGCAGTTACCTTAAATGGTGTGGACAAATCTTA 1162
Qy 961 TTAATTTAACCAACAGCGTAATTAATGGCTCACAGTGGTTACCTTTGCTAAACCTCAA 1020
Db 1163 TTAATTTAACCAACAGCGTAATTAATGGCTCACAGTGGTTACCTTTGCTAAACCTCAA 1222
Qy 1021 GTTAAGACGTTGTTTATCAAGGCTGGCACCAATTCGTCCTTAGTAAACAGAAACAGAT 1080
Db 1223 GTTAAGACGTTGTTTATCAAGGCTGGCACCAATTCGTCCTTAGTAAACAGAAACAGAT 1282
Qy 1081 GTGACTTTGATATTGATGACCAACAAATTTGAAAACAGGCCCTTACATCCAGTTTAGGG 1140
Db 1283 GTGACTTTGATATTGATGACCAACAAATTTGAAAACAGGCCCTTACATCCAGTTTAGGG 1342
Qy 1141 ATCTTGATCAATGGGCAACCCAAATTTGAAAATATCTTGGAAAGAGAACTCTTCAG 1200
Db 1343 ATCTTGATCAATGGGCAACCCAAATTTGAAAATATCTTGGAAAGAGAACTCTTCAG 1402
Qy 1201 TTTGATGTCCAAAGTTGCGAATCTACTGTGACCCAGAACAGAAACCCGGGAGACAGCA 1260
Db 1403 TTTGATGTCCAAAGTTGCGAATCTACTGTGACCCAGAACAGAAACCCGGGAGACAGCA 1462
Qy 1261 TGTGAGATTCCTGGATTT-----TGCCTTAATGCTCCAGTGTAGTTCRACT 1311
Db 1463 TGTGAGATTCCTGGATTTAATGGAGAGTGCCTTAATGCTCCAGTGTAGTTCRACT 1522
Qy 1312 CCAGCTCCCTGTATTTGCTCCGGGAAACCCAGGACTTCAAGGCCCCCAAGGTGACCCCT 1371
Db 1523 CCAGCTCCCTGTATTTGCTCCGGGAAACCCAGGACTTCAAGGCCCCCAAGGTGACCCCT 1582
Qy 1372 GSNCTGCTGGNAACCTGGCTACCTCGACAGCACTGTGTCAAGTGTGAAGCTGGATAT 1431
Db 1583 GSNCTGCTGGNAACCTGGCTACCTCGACAGCACTGTGTCAAGTGTGAAGCTGGATAT 1642
Qy 1432 CAGGGAAATTCAGGGACACCCAGGTGTTCCAGGATCTCCAGGAAATCAAGGAGCTCGAGGA 1491
Db 1643 CAGGGAAATTCAGGGACACCCAGGTGTTCCAGGATCTCCAGGAAATCAAGGAGCTCGAGGA 1722
Qy 1492 CTACAGGTTACAAAGGAGAACCCAGGGCGAGATGGTGACAGGGTGATCGTGGACTTCCT 1551
Db 1703 CTACAGGTTACAAAGGAGAACCCAGGGCGAGATGGTGACAGGGTGATCGTGGACTTCCT 1762
Qy 1552 GGTTCCTCGGGCTTCATGGCATGCGCAGGATCAAGGGTGAATGGGTGCCAAAGGAGAC 1611
Db 1763 GGTTCCTCGGGCTTCATGGCATGCGCAGGATCAAGGGTGAATGGGTGCCAAAGGAGAC 1822
Qy 1612 AAGAGTACCTGGATTTTATGGCAAAAAGGTGCAAAAGGTGAAAAGGGGAAATGCTGGC 1671
Db 1823 AAGAGTACCTGGATTTTATGGCAAAAAGGTGCAAAAGGTGAAAAGGGGAAATGCTGGC 1882
Qy 1672 TTCCTGGCTCCCTGGACTGCTGAGAACCCAGGAGACATGGAAGAGATGGAATTAATG 1731
Db 1883 TTCCTGGCTCCCTGGACTGCTGAGAACCCAGGAGACATGGAAGAGATGGAATTAATG 1942
Qy 1732 GGTAGTCCCGTTTCAAGGAGAGAGAGTCCCTGGTGTCTCCGGGGCAGGATGGAACA 1791
Db 1943 GGTAGTCCCGTTTCAAGGAGAGAGAGTCCCTGGTGTCTCCGGGGCAGGATGGAACA 2002
Qy 1792 CGGGGAGAGCTCGAATCCAGGATTTCTGGAACCCAGGATTAATGGGCCAAAGGGA 1851

RESULT 4
HSM801597

Db 2003 CGGCGAGAGCCTGGATCCAGGATTTCTGGAACCCAGAGATTAAATGGGCCCAAGGGA 2062
Qy 1852 GAAATTTGGGCTCTCAGGACAGCAAGGAAAAAAGGAGCCCGAGGATGCTGTTTAATG 1911
Db 2063 GAAATTTGGGCTCTCAGGACAGCAAGGAAAAAAGGAGCCCGAGGATGCTGTTTAATG 2122
Qy 1912 GGAAGCAATGCTCACCAGGCGCAGCTTGGAAACCCGGGATCTAAGGGAACAAAGGTGA 1971
Db 2123 GGAAGCAATGCTCACCAGGCGCAGCTTGGAAACCCGGGATCTAAGGGAACAAAGGTGA 2182
Qy 1972 CTTGGAATTCAGGATCCTCGGGGCTTCTGGGCTCAAGGGAGAACCCAGGAGCAACGGGT 2031
Db 2183 CTTGGAATTCAGGATCCTCGGGGCTTCTGGGCTCAAGGGAGAACCCAGGAGCAACGGGT 2242
Qy 2032 TCCCAGAGAGAACCCAGGATACATGGGTTTACCCGGATTCAGGAAAAAAGGGGAGCAAA 2091
Db 2243 TCCCAGAGAGAACCCAGGATACATGGGTTTACCCGGATTCAGGAAAAAAGGGGAGCAAA 2302
Qy 2092 GGAATTCAGGATGAAAAAGGTATTCAAGGCTCAAAAGGGAGAAATGGAACACAGGGAAT 2151
Db 2303 GGAATTCAGGATGAAAAAGGTATTCAAGGCTCAAAAGGGAGAAATGGAACACAGGGAAT 2362
Qy 2152 CCAGGCGAACAGGGGAATTCAGGCGAATCATGCTGCAAAAGGAGAGAGAGTGAAGGGA 2211
Db 2363 CCAGGCGAACAGGGGAATTCAGGCGAATCATGCTGCAAAAGGAGAGAGAGTGAAGGGA 2422
Qy 2212 GAACTTGTGTCCAGGCTGCAATTTGGATCAAAAGGAGAACTCTGGGTTGGATGGCTGATG 2271
Db 2423 GAACTTGTGTCCAGGCTGCAATTTGGATCAAAAGGAGAACTCTGGGTTGGATGGCTGATG 2482
Qy 2272 GGGCCCGGAGTCTTAAGGGCAACCTGGGGATCCAGGCTCCTCAGGAGACCCCGAGTTTG 2331
Db 2483 GGGCCCGGAGTCTTAAGGGCAACCTGGGGATCCAGGCTCCTCAGGAGACCCCGAGTTTG 2542
Qy 2332 GATGGGAAGCCCGGAGAGAGTTCAGAACAAATTTATTCAGAAAGTTTGCACAGATGTA 2391
Db 2543 GATGGGAAGCCCGGAGAGAGTTCAGAACAAATTTATTCAGAAAGTTTGCACAGATGTA 2602
Qy 2392 ATAAGAGCCCGAGTACCAAGTCTTACTTCAGAGTGGAGAAATTAAGAAATTTGATGTC 2451
Db 2603 ATAAGAGCCCGAGTACCAAGTCTTACTTCAGAGTGGAGAAATTAAGAAATTTGATGTC 2662
Qy 2452 CTGTCCTCAGATGGCTCCCGGATTCCTCGGCGACCTGGTCCGATAGGCCAGAGGT 2511
Db 2663 CTGTCCTCAGATGGCTCCCGGATTCCTCGGCGACCTGGTCCGATAGGCCAGAGGT 2722
Qy 2512 CCAGAGGATTAACCTGGTTTGCAGGAGAGATGGTCTCTCGGATTTAGTGGGTGCTCCCT 2571
Db 2723 CCAGAGGATTAACCTGGTTTGCAGGAGAGATGGTCTCTCGGATTTAGTGGGTGCTCCCT 2782
Qy 2572 GSNCTCAGGCTGTCCAGAGGATTAAGGCGCTACAGGAGAAATGGGAAAAAGGGAGC 2631
Db 2783 GSNCTCAGGCTGTCCAGAGGATTAAGGCGCTACAGGAGAAATGGGAAAAAGGGAGC 2842
Qy 2632 CAAGGTTTGGGTATCTCGGAGAACCAAGTCTCTCGTCCCGAGGCTCCAGAGGGCCCT 2691
Db 2843 CAAGGTTTGGGTATCTCGGAGAACCAAGTCTCTCGTCCCGAGGCTCCAGAGGGCCCT 2902
Qy 2692 CTTGGAATAAGCAAGAGAGTCTCCAGAGAGACCCAGGCTCTCTGTCGCAAGATGAGAGC 2751
Db 2903 CTTGGAATAAGCAAGAGAGTCTCCAGAGAGACCCAGGCTCTCTGTCGCAAGATGAGAGC 2962
Qy 2752 CATGGAACCTTGGATTCAGGCGAACCCAGGCGCCCGAGGATCTCGACCCATCACTA 2811
Db 2963 CATGGAACCTTGGATTCAGGCGAACCCAGGCGCCCGAGGATCTCGACCCATCACTA 3022
Qy 2812 TGTCTTAGTGTAAATTCAGAGAGATTCCTGTCAGAAAGSACCAAACTATTAG 2865
Db 3023 TGTCTTAGTGTAAATTCAGAGAGATTCCTGTCAGAAAGSACCAAACTATTAG 3076

<p> Query Match 99.2%; Score 2841.2; DB 9; Length 4160; Best Local Similarity 99.6%; Pred. No. 0; Mismatches 2862; Conservative 0; Mismatches 3; Indels 9; Gaps 1; </p>									
QY	1	ATGGCTCACTATATACATTTCTGCAATGTTTGGTGGCTCTCTTTCAGAAATCTGTG	60						
DB	203	ATGGCTCACTATATACATTTCTGCAATGTTTGGTGGCTCTCTTTCAGAAATCTGTG	262						
QY	61	TTAGCTGAGATGGGAGATGATCAAGTTCCTGCTACTGCTCCGACAGATTTAGTTTC	120						
DB	263	TTAGCTGAGATGGGAGATGATCAAGTTCCTGCTACTGCTCCGACAGATTTAGTTTC	322						
QY	121	ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAATTTGAAATAGTGAAGAGTGGCTT	180						
DB	323	ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAATTTGAAATAGTGAAGAGTGGCTT	382						
QY	181	GTCAATATCAAAAACATTTGACATAGGCGCGAAGTTTATTCAGTTGGAGTGGTCAA	240						
DB	383	GTCAATATCAAAAACATTTGACATAGGCGCGAAGTTTATTCAGTTGGAGTGGTCAA	442						
QY	241	TATAGTGAATACCTGCTGTGAGATTTCTCTCGAAGCTATGATTCAGGAGAACATTTG	300						
DB	443	TATAGTGAATACCTGCTGTGAGATTTCTCTCGAAGCTATGATTCAGGAGAACATTTG	502						
QY	301	ACGGCAGAGTGAATTCATCTTACTTAGGAGGAAACAAAGACAGGAGAGCCATC	360						
DB	503	ACGGCAGAGTGAATTCATCTTACTTAGGAGGAAACAAAGACAGGAGAGCCATC	562						
QY	361	CAGTTTGGCTCGATTTACCTTTTGGCAAGTCTCAGATTTCTGACTAAGATAGCAGTG	420						
DB	563	CAGTTTGGCTCGATTTACCTTTTGGCAAGTCTCAGATTTCTGACTAAGATAGCAGTG	622						
QY	421	GTACTTACGATGGCAATTCCTCAAGATGAGCTCAAGATGAGCTCAAGCAGCAGAGAT	480						
DB	623	GTACTTACGATGGCAATTCCTCAAGATGAGCTCAAGATGAGCTCAAGCAGCAGAGAT	682						
QY	481	AGTAAGATAACATTTATTTGCTATTTGGTTTGGTTTTCAGAAAACAGAGATGCCAATAGA	540						
DB	683	AGTAAGATAACATTTATTTGCTATTTGGTTTGGTTTTCAGAAAACAGAGATGCCAATAGA	742						
QY	541	GCATTGCCAACAGCTTTGCTCTACTTATGTTTATTTATGTCGAAAGACTATATTCGAATA	600						
DB	743	GCTATTGCCAACAGCTTTGCTCTACTTATGTTTATTTATGTCGAAAGACTATATTCGAATA	802						
QY	601	TCCAAATAGGAGATGATGAGCAGAACTTTGTGAGAACTCTGTCTGTCACACAGA	660						
DB	803	TCCAAATAGGAGATGATGAGCAGAACTTTGTGAGAACTCTGTCTGTCACACAGA	862						
QY	661	ATTCAGTGGCAGCTCGTGAATGAAAGGGATTTGATATTTCTTTTAGGTTTAGATGTAAT	720						
DB	863	ATTCAGTGGCAGCTCGTGAATGAAAGGGATTTGATATTTCTTTTAGGTTTAGATGTAAT	922						
QY	721	AAAAAGTTTAAAGAAAGATACAGCTTTTCCAAAAGAGATTAAGAGGATATGAAGTACA	780						
DB	923	AAAAAGTTTAAAGAAAGATACAGCTTTTCCAAAAGAGATTAAGAGGATATGAAGTACA	982						

QY	781	TCAAAAGTGTATTTATCAGAACTCACAAGCAATGTTTCCCAAGAGGTCTTCTCCATCA	840						
DB	983	TCAAAAGTGTATTTATCAGAACTCACAAGCAATGTTTCCCAAGAGGTCTTCTCCATCA	1042						
QY	841	TATGATATTTGCTCTACTCAAAAGATTTAAAGTCAAGAAAATTTGGGATTTATGAGAA	900						
DB	1043	TATGATATTTGCTCTACTCAAAAGATTTAAAGTCAAGAAAATTTGGGATTTATGAGAA	1102						
QY	901	TTAACTATTTGATGGAAGGCCCAAAATAGCAGTTACCTTAAATGCTGGGACAAATCTTA	960						
DB	1103	TTAACTATTTGATGGAAGGCCCAAAATAGCAGTTACCTTAAATGCTGGGACAAATCTTA	1162						
QY	961	TTATTATCAAGCAAGCAGCTAATTAAAGCTCAAGTGGTACCTTTGCTAACCCCTCAA	1020						
DB	1163	TTATTATCAAGCAAGCAGCTAATTAAAGCTCAAGTGGTACCTTTGCTAACCCCTCAA	1222						
QY	1021	GTAAAGAGCTTTGTTGATCAAGGCTGGCACCACAAATTCCTCTTAGTAAACAGAACAGAT	1080						
DB	1223	GTAAAGAGCTTTGTTGATCAAGGCTGGCACCACAAATTCCTCTTAGTAAACAGAACAGAT	1282						
QY	1081	GTGACTTTGATATTTGATGACCAAAATTTGAAAACAGCCCTTACATCCAGTTTATGGG	1140						
DB	1283	GTGACTTTGATATTTGATGACCAAAATTTGAAAACAGCCCTTACATCCAGTTTATGGG	1342						
QY	1141	ATCTTGATCAATGGGCAAAACCCAAATTTGAAAATATTTCTGAAAAGAGAAACTTTTCA	1200						
DB	1343	ATCTTGATCAATGGGCAAAACCCAAATTTGAAAATATTTCTGAAAAGAGAAACTTTTCA	1402						
QY	1201	TTTGATGTCAAAAGATTGCGAATCTTACTGTGACCCAGAACAGAACACCCGGGAGACAG	1260						
DB	1403	TTTGATGTCAAAAGATTGCGAATCTTACTGTGACCCAGAACAGAACACCCGGGAGACAG	1462						
QY	1261	TGTGAGATTCCTGGGATTT-----TGCCCTTAATGGTCCACGTGATGTAGGTTCAACT	1311						
DB	1463	TGTGAGATTCCTGGGATTTAAATGGAGATGTCCTTAAATGGTCCACGTGATGTAGGTTCAACT	1522						
QY	1312	CCAGCTCCCTGTATTTGTCTCCGGGAAAACCCAGACTTCAAGGCCCCCAAGGTGACCTT	1371						
DB	1523	CCAGCTCCCTGTATTTGTCTCCGGGAAAACCCAGACTTCAAGGCCCCCAAGGTGACCTT	1582						
QY	1372	GGACTGCTGGGAAACCTTACCTGACCAACTGCTGAGATGCTGAGCTGAGAT	1431						
DB	1583	GGACTGCTGGGAAACCTTACCTGACCAACTGCTGAGATGCTGAGCTGAGAT	1642						
QY	1432	CAGGAAATTCAGGAGACACAGGCTTCCAGGATCTCCAGGAATACAAGAGCTCGAGGA	1491						
DB	1643	CAGGAAATTCAGGAGACACAGGCTTCCAGGAATACAAGAGCTCGAGGA	1702						
QY	1492	CTACCAAGTTTCAAAAGGAGAACCCAGGCGAGATGGTGAACAGGCTGATCGGACTTCCT	1551						
DB	1703	CTACCAAGTTTCAAAAGGAGAACCCAGGCGAGATGGTGAACAGGCTGATCGGACTTCCT	1762						
QY	1552	GGTTTTCCTGGGCTTCATGGATGCCAGATCAAAAGGTGAAATGGGTGCCAAAGGAGAC	1611						
DB	1763	GGTTTTCCTGGGCTTCATGGATGCCAGATCAAAAGGTGAAATGGGTGCCAAAGGAGAC	1822						
QY	1612	AAAGGATCACCTGGATTTTATGGGAAAAGGTGCAAAAGGTGAAAAGGGGATGCTGGC	1671						
DB	1823	AAAGGATCACCTGGATTTTATGGGAAAAGGTGCAAAAGGTGAAAAGGGGATGCTGGC	1882						
QY	1672	TTCCCTGGGCTCCCTGGACCTGCTGGAGAACCCAGGAGACATGAAAGGATGGAATTAATG	1731						
DB	1883	TTCCCTGGGCTCCCTGGACCTGCTGGAGAACCCAGGAGACATGAAAGGATGGAATTAATG	1842						
QY	1732	GGTAGTCCCGGTTTCAAGGGAGAACAGGATCCCTGCTGCTCCGGGCGAGATGGAACA	1791						
DB	1943	GGTAGTCCCGGTTTCAAGGGAGAACAGGATCCCTGCTGCTCCGGGCGAGATGGAACA	2002						
QY	1792	CGGGAGAGCCTCGAATTCAGGATTTCTCTGAAAACCCAGGATTAATGGGCGCAAAAGGA	1851						
DB	2003	CGGGAGAGCCTCGAATTCAGGATTTCTCTGAAAACCCAGGATTAATGGGCGCAAAAGGA	2062						

BASE COUNT 1270 a 796 c 1026 g 1068 t

ORIGIN

QY	1321	TGTATTGTCTCCGGGAAACCCAGGACTTCAAGGCCCCAAAGGTGACCTCTGGACTGCCT	1380
Db	1483	TGTATTGTCTCCGGGAAACCCAGGACTTCAAGGCCCCAAAGGTGACCTCTGGACTGCCT	1542
QY	1381	GGGAACCTGGCTTACCTCGACAACTCGTCAAGATGGTAAGCCCTGGATATCAGGGAAT	1440
Db	1543	GGGAACCTGGCTTACCTCGACAACTCGTCAAGATGGTAAGCCCTGGATATCAGGGAAT	1602
QY	1441	GCAGGACACCCAGGTGTTCCAGGATCTCCAGGAATACAGGAGCTCGAGGACTACAGGT	1500
Db	1603	GCAGGACACCCAGGTGTTCCAGGATCTCCAGGAATACAGGAGCTCGAGGACTACAGGT	1562
QY	1501	TACAAAGGAGAACCCAGGCGAGATGCTGACAAAGGTGATCGTGACTTCTCTGGTTTCT	1560
Db	1663	TACAAAGGAGAACCCAGGCGAGATGCTGACAAAGGTGATCGTGACTTCTCTGGTTTCT	1722
QY	1561	GGGCTTCATGGCATGCCAGATCAAAAGGTGAAATGGGTGCGCAAGGAGACAAAGGATCA	1620
Db	1723	GGGCTTCATGGCATGCCAGATCAAAAGGTGAAATGGGTGCGCAAGGAGACAAAGGATCA	1782
QY	1621	CCTGGATTATGGCAAAAGGTCGCAAGGTGAAAGGGGCAATGCTGGCTTCCCTGGC	1680
Db	1783	CCTGGATTATGGCAAAAGGTCGCAAGGTGAAAGGGGCAATGCTGGCTTCCCTGGC	1842
QY	1681	CTCCTGGACCTGCTGGAGAACCCAGGAGACATGGAAAGGATGGATTAAATGGGTAGTCCC	1740
Db	1843	CTCCTGGACCTGCTGGAGAACCCAGGAGACATGGAAAGGATGGATTAAATGGGTAGTCCC	1902
QY	1741	GGTTTCAAGGAGAGCAGGATCCCTGCTGCTCGGGCGAGATGGGAACCGGGGAGAG	1800
Db	1903	GGTTTCAAGGAGAGCAGGATCCCTGCTGCTCGGGCGAGATGGGAACCGGGGAGAG	1862
QY	1801	CCTGGAATCCAGGATTTCTCTGGAAACCGAGGATTAATGGGCCAAAGGGAGAAATGGG	1860
Db	1963	CCTGGAATCCAGGATTTCTCTGGAAACCGAGGATTAATGGGCCAAAGGGAGAAATGGG	2022
QY	1861	CTCCAGGACAGGAGGAAAAAGAGGCCCGCAGGATGCTGGTTTAATGGGAGCAAT	1920
Db	2023	CTCCAGGACAGGAGGAAAAAGAGGCCCGCAGGATGCTGGTTTAATGGGAGCAAT	2082
QY	1921	GGCTCACCGGCGAGCTCGAACACCGGGATCTAAGGGAAGCAAAAGGTGAACCTGGGAAT	1980
Db	2083	GGCTCACCGGCGAGCTCGAACACCGGGATCTAAGGGAAGCAAAAGGTGAACCTGGGAAT	2142
QY	1981	CAAGGATGCTGGGCTCTCGGCTCAAGGGAGAACCCAGAGCAACCGGTTCCCGAGGA	2040
Db	2143	CAAGGATGCTGGGCTCTCGGCTCAAGGGAGAACCCAGAGCAACCGGTTCCCGAGGA	2202
QY	2041	GAACCCAGGATACATGGTGTACCCGGGATTCAGGAAAAAGGGGCAAAAGGAAATCAA	2100
Db	2203	GAACCCAGGATACATGGTGTACCCGGGATTCAGGAAAAAGGGGCAAAAGGAAATCAA	2262
QY	2101	GGTGAAGGATTCAGGCTCAAGGGAGAAAAATGGAAGCAGGGAATTCAGGGGCA	2160
Db	2263	GGTGAAGGATTCAGGCTCAAGGGAGAAAAATGGAAGCAGGGAATTCAGGGGCA	2322
QY	2161	CAGGGAATTCAGGCTCATGTGTCAAAAGCAGAGAGAGGTGAAGGAGAAACCTGGT	2220
Db	2323	CAGGGAATTCAGGCTCATGTGTCAAAAGCAGAGAGAGGTGAAGGAGAAACCTGGT	2382
QY	2221	GTCCAGGTCCTCATGATCAAAAGGAGAAATCTGGGTGGATGGCTTGATGGGCCCGCA	2280
Db	2383	GTCCAGGTCCTCATGATCAAAAGGAGAAATCTGGGTGGATGGCTTGATGGGCCCGCA	2442
QY	2281	GGTCTTAAGGGCAACTCGGATCCAGGTCCTCAGGGACCCCAAGGTTTGGATGGGAAG	2340
Db	2443	GGTCTTAAGGGCAACTCGGATCCAGGTCCTCAGGGACCCCAAGGTTTGGATGGGAAG	2502
QY	2341	CCCGAAGAGAGTTTTCAGAACAAATTTATTCACAAAGTTTGCACAGATGTAATAAGAGCC	2400
Db	2503	CCCGAAGAGAGTTTTCAGAACAAATTTATTCACAAAGTTTGCACAGATGTAATAAGAGCC	2562
QY	2401	CAGTACCAGTCTTACTTCAGAGTGAAGAAATAGNAATTTGATCATCTGCTGCCAA	2460
Db	2563	CAGTACCAGTCTTACTTCAGAGTGAAGAAATAGNAATTTGATCATCTGCTGCCAA	2622
QY	2461	CATGGCTCCCCGGGTATTCCTGGGCCACCTGGTCCGATAGGCCAGAGGGTCCCAAGGA	2520
Db	2623	CATGGCTCCCCGGGTATTCCTGGGCCACCTGGTCCGATAGGCCAGAGGGTCCCAAGGA	2682
QY	2521	TTACTTGGTTTCCAGGAAGAGATGGTTCCTGGATTTAGTGGGTCTCCTGGACGTCA	2580
Db	2683	TTACTTGGTTTCCAGGAAGAGATGGTTCCTGGATTTAGTGGGTCTCCTGGACGTCA	2742
QY	2581	GGTGTCAAGGATTTAAAGGCTTACCAAGGAAGAAATGGGAAAAAGGGAGCCCAAGGGTTT	2640
Db	2743	GGTGTCAAGGATTTAAAGGCTTACCAAGGAAGAAATGGGAAAAAGGGAGCCCAAGGGTTT	2802
QY	2641	GGGTATCTCTGAGAACCAAGGCTCTCTGGTCCCTCCAGGTCCTCAGAGGGCCCTCTGGAATA	2700
Db	2803	GGGTATCTCTGAGAACCAAGGCTCTCTGGTCCCTCCAGGTCCTCAGAGGGCCCTCTGGAATA	2862
QY	2701	AGCAAGAGAGGTCTCTCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGACCATGGAAA	2760
Db	2863	AGCAAGAGAGGTCTCTCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGACCATGGAAA	2922
QY	2761	CCTGGAATCCAGGCGAACCCAGGCCCCCGAGGATCTCGAACCATCATCTTTAGT	2820
Db	2923	CCTGGAATCCAGGCGAACCCAGGCCCCCGAGGATCTCGAACCATCATCTTTAGT	2982
QY	2821	GTAATTCGCAAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG	2865
Db	2983	GTAATTCGCAAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG	3027
RESULT 2			
AF414088			
LOCUS	Homo sapiens collagen XXI (COL21A1) mRNA, complete cds.	4160 bp	mRNA linear PRI 16-NOV-2001
DEFINITION			
ACCESSION	AF414088		
VERSION	AF414088.1	GI:15593269	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM			
REFERENCE			
AUTHORS	Fitzgerald, J. and Bateman, J.F.		
TITLE	A new FACIT of the collagen family: COL21A1		
JOURNAL	FEBS Lett. 505 (2), 275-280 (2001)		
MEDLINE	21450665		
PUBMED	11566190		
REFERENCE			
AUTHORS	Fitzgerald, J. and Bateman, J.F.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-AUG-2001) Cell and Matrix Biology, Murdoch Children's Research Institute, Flemington Rd, Parkville, Vic 3052, Australia		
FEATURES			
source			
	1..4160		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="6"		
	/map="6p12-11"		
	1..4160		
	/gene="COL21A1"		
	203..3076		
	/gene="COL21A1"		
	/note="FACIT collagen; similar to protein encoded by Genbank Accession Number AL136624"		
	/codon_start=1		
	/product="collagen XXI"		
	/protein_id="AA0227.1"		
	/db_xref="GI:15593270"		
	/translation="MAHYITFLCMVLVLLIQLNSVLAEDGEVRSRSTAPDILVFLDGLSYSGPENFEIVRKMLNITNFIDGPKFIQGVQYSDYPLVLEPLGSGTSGEHLTA		
gene			
CDS			

collagen (COL21A1) gene		
JOURNAL	Genomics 79 (3), 395-401 (2002)	
MEDLINE	21853310	
PUBMED	11863369	
REFERENCE	2 (bases 1 to 4122)	
AUTHORS	Chou,M.-Y.	
TITLE	Cloning and identification of a novel human alpha 1 chain-like	
JOURNAL	collagen (COLA1L) gene, a new member of the FACIT family	
REFERENCE	Unpublished	
AUTHORS	3 (bases 1 to 4122)	
TITLE	Chou,M.-Y.	
JOURNAL	Submitted (20-DEC-2000) Biomedical Engineering Center, Industrial	
FEATURES	Technology Research Institute, 195-85, Sec. 4, Chung Hsing Rd.,	
source	Chutung, Hsinchu 310, Taiwan	
	Location/Qualifiers	
	1. .4122	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/chromosome="6"	
	/map="6p12.3-p11.2"	
	1. .4122	
	/gene="COLA1L"	
	163. .3027	
	/gene="COLA1L"	
	/codon_start=1	
	/product="alpha 1 chain-like collagen COLA1L precursor"	
	/protein_id="AAU50033.1"	
	/db_xref="GI:11974510"	
	/translation="MAHYITFLCMLVLLQNSVLABGEVRSRCRTAPTDLVFLDGL	
	SYSGENPEIVNKLWNIQNFIDLPKFIQGVVQYSDYVLEIPLSYDSGEHLTA	
	AVESILYFGNTGKAIQFALDYPKRSRLTKIAVLITDGLQSDQVDAQAARD	
	SKITLPAIGVGSDEDAELRAITANKSPSYVYVEDYIAISKIREVMKOLCESVCP	
	TRIPVAARDRGRDIIILGLDVNKKVKRIQLSPKIKGVENTSKVDLSELTSNVFPEG	
	LPSPYVYSTQREKVKIWDLPRLITIDRPPOLAVTLNGVDKILLFTTTSVINGSOVV	
	TFANPQVTLDFEGWQIRLLVTEQDVLITIDQIENKPLHPVLGILLINQGTQIGKY	
	SGKEITVDFQVQLRIYCDPEQNNRETAICEIPGFLNGPSDVSTPAPCIPCPKPGKL	
	QGPGLDPLFNPNGYPOQDQKPGYGIAGTGPVPGSPGIQAGRLGPKYKGFPRD	
	GDKGDRGLPGLFGLHMPGSKGEMAKGDKGSPFYKKGAKGKNGAGFPGLPGPAG	
	EPGRHGRDGLMGSPFKFAGSGPAGPDGTGEPIPGFPGNRLMGOKGEIGPPGP	
	QKKGKAGPMLGMSNGSPQPCPTGSKSGKEPGILOGMPGASGLKSPGATGSPGP	
	GYMGLPGIQCKKDGNOKEFGILOGKNGRQGIPOGGIOGHGKGERGEKGP	
	VRGALSGKSGVDGLMGHPAGPKQPGDPPQPGPLDQKQFRESQFIRQVCTEVI	
	RAQVPLIQSGRLNCDHCLSGHSGFIPGPPTGPFGPGLPGLPGRDGVPLGLVGV	
	PGRPVRLGLQKFLPGRNGEKGSGFGYGPSPGPPGPPGPGIISKEGPPGDPGLPGK	
	DGDHKGPGIQGQFPFGICDPSLCFSFVIARRDPFRKPNY"	
	163. .228	
	/gene="COLA1L"	
	4072. .4077	
	/gene="COLA1L"	
BASE COUNT	1280 a 769 g 996 g 1077 t	
ORIGIN		
Query Match	100.0%; Score 2865; DB 9; Length 4122;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 2865; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1 ATGGCTCACTATATACATTCTCTGCATGGTTTGGTCTGCTCTTTCAGAAATCTGTG 60	
Db	163 ATGGCTCACTATATACATTCTCTGCATGGTTTGGTCTGCTCTTTCAGAAATCTGTG 222	
Qy	61 TTAGCTGAAGTGGGGAAGTAAGATCAAGTTGTGCTACTGCTCCGACAGATTAGTTTC 120	
Db	223 TTAGCTGAAGTGGGGAAGTAAGATCAAGTTGTGCTACTGCTCCGACAGATTAGTTTC 282	
Qy	121 ATCTTAGATGCTCTTAGTGTGGGCCGAGAAAACCTTTGAAATAGTGAAGAAGTGGCTT 180	
Db	283 ATCTTAGATGCTCTTAGTGTGGGCCGAGAAAACCTTTGAAATAGTGAAGAAGTGGCTT 342	
Qy	181 GTCAATATCACAAAANAACCTTTGACATAGGCGCGAAGTTTATTCAAGTTGGAGTGGTTCAA 240	
Db	343 GTCAATATCACAAAANAACCTTTGACATAGGCGCGAAGTTTATTCAAGTTGGAGTGGTTCAA 402	

241	TATAGTGA	ACTACCC	TGCTCG	GATTCCT	CTCGA	AGCTAT	GATTC	GAGAAC	ATTTC	300
403	TATAGTGA	CTACCC	TGCTCG	GATTCCT	CTCGA	AGCTAT	GATTC	GAGAAC	ATTTC	462
301	ACGGCAG	CAGTGG	GAATCC	ATACT	CTACT	TAGG	AGGA	CAACA	AGAGC	360
463	ACGGCAG	CAGTGG	GAATCC	ATACT	CTACT	TAGG	AGGA	CAACA	AGAGC	522
361	CAGTTTGC	GCTCGA	TACCT	TTTTT	GCCA	AGTC	CTC	CAGAT	TTCT	420
523	CAGTTTGC	GCTCGA	TACCT	TTTTT	GCCA	AGTC	CTC	CAGAT	TTCT	582
421	GTACTTAC	CGATG	CGAAT	CCCA	AGAT	GAG	TCG	AGG	ATG	480
583	GTACTTAC	CGATG	CGAAT	CCCA	AGAT	GAG	TCG	AGG	ATG	642
481	AGTAAAG	ATAAC	ATTAT	TGCT	TATT	GGTG	TTT	GGT	TTC	540
643	AGTAAAG	ATAAC	ATTAT	TGCT	TATT	GGTG	TTT	GGT	TTC	702
541	GCTATTGC	CAAC	AGCCT	CTCT	CTACT	TAT	TGT	TTT	ATG	600
703	GCTATTGC	CAAC	AGCCT	CTCT	CTACT	TAT	TGT	TTT	ATG	762
601	TCCAAA	ATAAG	GGAAG	TGAT	GAAG	CAG	AAAC	CTTT	TGT	660
763	TCCAAA	ATAAG	GGAAG	TGAT	GAAG	CAG	AAAC	CTTT	TGT	822
661	ATTCCAG	TGGAG	CTG	TGAT	GAA	GSG	GA	TTT	GAT	720
823	ATTCCAG	TGGAG	CTG	TGAT	GAA	GSG	GA	TTT	GAT	882
721	AAAAAG	TTTAA	AAAA	GAAT	CAC	AGT	TTT	CAC	AAAA	780
883	AAAAAG	TTTAA	AAAA	GAAT	CAC	AGT	TTT	CAC	AAAA	942
781	TCAAAG	TGAT	TTTAT	CAG	AACT	CAC	AAAG	TAAT	TTT	840
943	TCAAAG	TGAT	TTTAT	CAG	AACT	CAC	AAAG	TAAT	TTT	1002
841	TATGTA	TTTGT	CTCT	ACT	CAAA	AGT	TTT	AAAG	TTC	900
1003	TATGTA	TTTGT	CTCT	ACT	CAAA	AGT	TTT	AAAG	TTC	1062
901	TTAAC	TAT	TGAT	TGG	AGG	CC	CA	AAAT	TAG	960
1063	TTAAC	TAT	TGAT	TGG	AGG	CC	CA	AAAT	TAG	1122
961	TTATTT	TAC	AA	CAAC	CAG	CGT	TA	TTA	TG	1020
1123	TTATTT	TAC	AA	CAAC	CAG	CGT	TA	TTA	TG	1182
1021	GTTAAG	ACCT	GT	TTG	AT	G	AGG	CT	GG	1080
1183	GTTAAG	ACCT	GT	TTG	AT	G	AGG	CT	GG	1242
1081	GTGACT	TTT	CT	TAT	TAT	TAT	G	AT	T	1140
1243	GTGACT	TTT	CT	TAT	TAT	TAT	G	AT	T	1302
1141	ATCTT	GAT	CAAT	GGG	CAAA	CC	CAAT	TT	T	1200
1303	ATCTT	GAT	CAAT	GGG	CAAA	CC	CAAT	TT	T	1362
1201	TTTGT	GT	CT	CA	AAAG	CT	CT	CT	CT	1260
1363	TTTGT	GT	CT	CA	AAAG	CT	CT	CT	CT	1422
1261	TGT	GAG	AT	TCC	TGG	AT	TTT	TG	CT	1320
1423	TGT	GAG	AT	TCC	TGG	AT	TTT	TG	CT	1482

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 18:36:28 ; Search time 10133 Seconds
(without alignments)
11566.770 Million cell updates/sec

Title: US-09-996-611b-5
Perfect score: 2865
Sequence: 1 atgggtcactatattacatt.....gaaaggacaaactattag 2865

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.in.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2865	100.0	4122	9	AF330693	AF330693 Homo sapi
2	2841.2	99.2	4120	9	AF414088	AF414088 Homo sapi
3	2839.5	99.1	4160	6	AX086127	AX086127 Sequence
4	2839.6	99.1	4160	9	HSM801597	AL136624 Homo sapi
5	1648	36.6	2295	9	AK096444	AK096444 Homo sapi
6	974	34.0	1184	6	AX430410	AX430410 Sequence
7	972.4	33.9	1169	6	AX430281	AX430281 Sequence
8	908	31.7	1125	6	AX552238	AX552238 Sequence
9	553.5	19.3	110000	9	AF438327_2	Continuation (3 of
10	553.6	19.3	126501	9	HX708F5	AL031782 Human DNA
11	453.6	15.8	1414	6	AX721209	AX721209 Sequence
12	353.4	12.3	447	6	BD058397	BD058397 Secretd
13	303	10.6	3164	5	CHKCGCND	K01702 Chicken cho
14	295.4	10.3	2927	9	HSCOL9AL	X54412 Human mRNA
15	292.6	10.2	2359	3	AK115182	AK115182 Ciona int
16	289.6	10.1	2766	10	MJCOLA1IX	L12215 Mouse colla
17	289.6	10.1	3049	10	MUSNA411	D17511 Mus musculu
18	285	9.9	3619	9	BC042075	BC042075 Homo sapi
19	285	9.9	6352	9	AF406780	AF406780 Homo sapi
20	282.6	9.9	3178	5	AY113700	AY113700 Dario rer
21	279	9.7	2388	10	AF349718	AF349718 Mus muscu
22	276.4	9.6	3074	5	BC045244	BC045244 Xenopus l
23	272.4	9.5	2416	5	CHK3ACOL	M83179 Gallus gall
24	272.2	9.5	2901	5	GOCOLA3IX	X64712 G.gallus mr
25	268.2	9.4	4428	6	AX146430	AX146430 Sequence
26	263.2	9.2	3720	9	S57132	S57132 COL16A1-typ
27	262.8	9.2	4473	3	AF525468	AF525468 Hydra vul
28	262	9.1	3044	5	BC046861	BC046861 Xenopus l
29	261.6	9.1	5387	9	HJMCOL16A	Y92642 Homo sapien
30	261.6	9.1	5387	11	G28574	G28574 human STS S
31	259.8	9.1	4428	6	AX146422	AX146422 Sequence
32	259.8	9.1	4428	6	AX146424	AX146424 Sequence
33	259.6	9.1	3902	9	BC028178	BC028178 Homo sapi
34	251	8.8	3967	10	BC028248	BC028248 Mus muscu
35	251	8.8	4778	10	BC043089	BC043089 Mus muscu
36	251	8.8	4786	10	BC052398	BC052398 Mus muscu
37	250.2	8.7	2777	5	SGU07973	U07973 Gallus gall
38	248.8	8.7	1955	10	BC023940	BC023940 Mus muscu
39	248.8	8.7	5160	10	BC027766	BC027766 Mus muscu
40	248.2	8.7	2263	4	CFU07889	U07888 Canis fami
41	248	8.7	5646	10	AB041350	AB041350 Mus muscu
42	247	8.6	5909	3	NEMA2C4A	M67507 A.suum alph
43	246.4	8.6	4235	10	BC016479	BC016479 Mus muscu
44	245.4	8.6	2667	5	CHKCOLA1XI	M88593 Gallus gall
45	242.6	8.5	5460	6	AX302553	AX302553 Sequence

ALIGNMENTS

RESULT :
AF330693
LOCUS
DEFINITION Homo sapiens alpha 1 chain-like collagen COL1A1 precursor (COL1A1)
mRNA, complete cds.
ACCESSION AF330693
VERSION AF330693.1 GI:17974509
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4122)
AUTHORS Chou,M.Y. and Li,H.C.
TITLE Genomic organization and characterization of the human type XXI